



UNIVERZITET SINGIDUNUM
FAKULTET ZA INFORMATIKU I MENADŽMENT
Beograd, Bulevar Zorana Đinđića br. 44

DEPARTMAN ZA POSLEDIPLOMSKE STUDIJE
MASTER STUDIJE

- Master rad -

Optimizacija pretraživačke mašine

-MASTER STUDIJSKI PROGRAM-
SAVREMENE INFORMACIONE TEHNOLOGIJE



UNIVERZITET SINGIDUNUM
FAKULTET ZA INFORMATIKU I MENADŽMENT
Beograd, Bulevar Zorana Đinđića br. 44

Predmeti:

Studijsko istraživački rad (Prof. Dr Mladen Veinović)

Studijska praksa (Prof. dr Ranko Popović)

Mentor:

Dr Ranko Popović

Kandidat:

Nikola Zobenica

Br. indeksa: 410185/2009

Beograd, 2011.

SADRŽAJ

1. Uvod	5
2. Ideja i princip rada Web servisa.....	6
2.1 Karakteristike infrastrukture Web servisa	6
2.2 Princip rada Web servisa	7
2.4 Primeri upotrebe web servisa	10
2.5 Potencijali u domaćim uslovima	10
3. Razvoj web servisa korišćenjem .Net platforme.....	11
3.1 Uvod	11
3.2 Osnove windows communicatio foundation-a.....	11
3.3 Osnove WCF klase i pojmovi	13
3.4 Osnovne osobine endpointa	14
3.5 Donošenje odluke o korišćenju binding-a pri razvoju servisa	14
3.6 Razvoj aplikacija u .NET FRAMEWORKU 3.0	15
3.7 Napomene	16
4. Osnovni servisi.....	17
4.1 Javni servisi.....	17
4.2 Servisi za pretraživanje.....	19
4.3 Sigurnosni servisi	20
4.4 Sistemski servisi.....	20
4.5 Elektronska pošta	21
4.6 WORLD WIDE WEB	22
5. Registracija web servisa UDDi Registri.....	24
5.1 Pronalaženje web servisa na Internetu.....	24
5.2 Osnovne karakteristike UDDI-a.....	24
5.3 Protokoli za registre web servisa.....	27
5.4 Karakteristike UDDI-a sa aspekta arhitekture	27
5.5 UDDI Produkt zahtevi	27
5.6 Modelovanje podataka.....	28
5.7 Modelovanje četiri osnovna elementa UDDI-a.....	28
5.8 Strategija kategorizacije.....	28
5.9 UDDI kontrole i procesi	29
5.10 Validacija podataka	30
5.11 Sintaksna provera	30
5.12 Semantička provera	30
5.13 Modelovanje ne web servisa u UDDI servise	31
5.14 UDDI API	31
5.15 Primer primene UDDI-a u realnom sistemu	31
6. Pretraživači	33
7. Teoretske osnove pretrage Interneta	34
7.1 Pretraživanje WWW.....	34
7.2 Search engine	34
7.3 Napredno pretraživanje Interneta	34
8. Pretraživanje Interneta.....	36
8.1 Osnove pretraživanja	36
8.2 Strategija pretraživanja	36
8.3 Napredna pretraga	37
8.4 Pretraga slika.....	37
9. Razlika između web direktorijuma i search engine-a	38
10. Search engine	39
10.1 Web direktorijumi	39

10.2 Krolerski ili spider pretraživači.....	39
10.3 Meta pretraživači.....	40
10.4 Problemi pretrživača.....	40
11. Google	41
11.1 Arhitektura	41
11.2 Software	41
11.3 Redosled stranica.....	41
11.4 Veličina index-a	42
11.5 Usluga pretraživanja za portale	43
11.6 Google appliance.....	43
12. Domaći search engine.....	44
12.1 Problemi domaćih pretraživača	44
13. SEO - Search engine optimization	45
13.1 Web crawl-eri.....	46
13.2 SEO principi i tehnike	47
13.3 Google search engine friendly stranica.....	50
14. Optimizacija Web stranice „www.elzing.rs“	55
14.1 Uvod	55
14.2 Odabir ključnih reči.....	56
14.3 Description tag	57
14.4 XML Mapa sajta.....	59
14.5 Važnost linkova na sajtu	60
14.6 Prijava sajta na pretraživače.....	62
14.7 Google webmaster tools.....	63
14.8 Statistika za „www.elzing.rs“	67
15. Web pretraživač	72
15.1 Crawling.....	72
15.2 Indeksiranje.....	74
15.3 Rankiranje	76
15.4 Clustering.....	76
16. Zaključak	77
17. Literatura.....	78

1. Uvod

U okviru ovog master rada, biće prikazani osnovni aspekti razvoja Web servisa, Pretraživačke mašine kao i same optimizacije Web stranice na pretraživače. Takođe biće predstavljena komercijalna rešenja (Google), kao i razvoj sajta “www.elzing.rs” i njegova optimizacija. U prvom delu rada, opisan je razvoj i problematika vezana za Web servise. Drugi deo rada, je vezan za problematiku oko pretraživačkih mašina (arhitektura, razvoj, problematika). U trećem delu prikazano je praktično rešenje implementacije i optimizacije Web sajta “www.elzing.rs”.

Internet je danas heterogen i nestrukturiran (većina informacija prezentirana na Web-u nije strukturirana tako da omogući razmenu između kompjutera, nego je naglasak na preglednom prikazu za ljudsko oko. Do pojave XML Web servisa nije postojao široko prihvaćeni standard za razmenu podataka. Web servisi kao tehnologija predstavljaju logičan nastavak razvoja Interneta prema labavo povezanoj (loosely connected) strukturi, koja će pomalo zameniti karakterističnu heterogenost Interneta. Upravo zato kompanije svih veličina danas investiraju u razvoj Web servisa, jer oni prevladavaju nekompatibilnosti između postojećih formata podataka i računarskih sistema, te time snižuju troškove slanja podataka preko računarskih mreža.

Web servisi imaju svoj spektar primene koji je mnogo širi od onog koje pokrivaju Web aplikacije. Zainteresovane firme i pojedinci mogu se jednostavno pretplatiti na korišćenje web servisa, tako da ne moraju investirati u razvoj sopstvenih. Sve što je potrebno je da u svoje Web aplikacije implementiraju nekoliko redova programskog koda kako bi inkorporirali eksterne web servise.

XML web servisi (WS) mogu se primeniti u svim scenarijima gdje je potrebna razmena informacija među subjektima na mreži (Internetu ili intranet) - bilo u svrhu obavljanja poslovnih transakcija ili jednostvne razmene vesti među Web portalima. Da bi to bilo moguće, Web servisi moraju biti potpuno nezavisni od sistema ili programskog jezika implementacije softvera

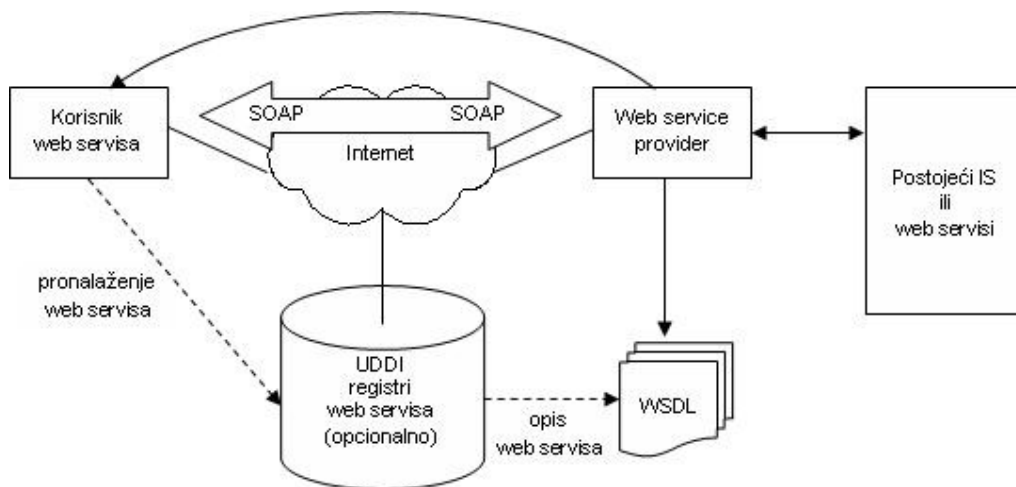
2. Ideja i princip rada Web servisa

Pod Web servisom se podrazumeva bilo koja softverska aplikacija koja je dostupna preko mreže (Internet ili intranet) i koja koristi standardizirani XML sistem poruka za razmenu podataka. Razmena funkcioniše tako da korisnik poziva Web servis slanjem upita u XML formatu, a poslužitelj u istom formatu vraća rezultat obrade.

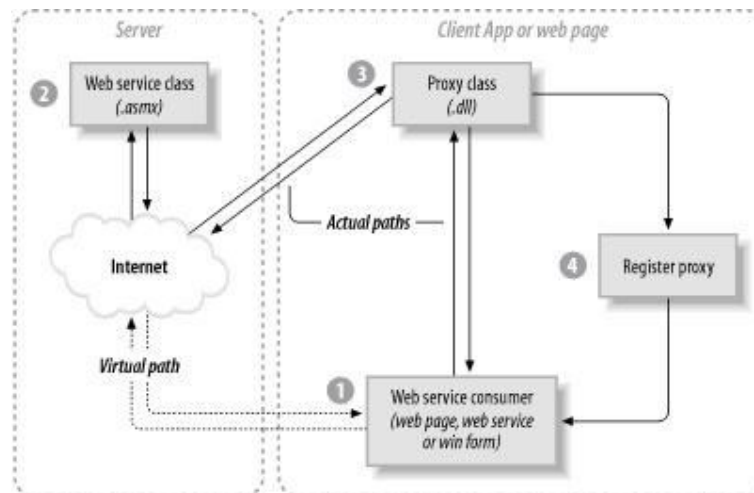
Budući se sva komunikacija odvija uz pomoć XML-a, Web servisi nisu vezani ni za jedan operativni sistem ili razvojno okruženje, čime se zaobilazi nekompatibilnosti među postojećim softverima.

2.1 Karakteristike infrastrukture Web servisa

- Web servis poslužitelj i korisnik Web servisa su spojeni na Internet, te mogu komunicirati međusobno i s drugim servisima
- način komunikacije mora odgovarati nekom široko prihvaćenom standardu - obično su to SOAP poruke, a mogu i biti HTTP-GET ili HTTP-POST zahtevi
- rad Web servisa ne zavisi o operacijskom sistemu, objektnom modelu niti o programskom jeziku, sve dok se Web servisi i korisničke aplikacije pridržavaju otvorenih standarda



2.2 Princip rada Web servisa



Korisnik Web servisa(1) šalje upit prema poslužitelju(2). Korisnik ima privid da komunicira preko Interneta direktno sa Web servisom, a u stvari komunicira sa proxy klasom(3), koja je lokalna korisniku. Zadatak proxy-a je da obavi sve komunikacijske pojedinosti nižeg nivoa potrebne za slanje podataka preko Interneta do poslužitelja. Proxy isto tako prima rezultate i prezentira ih korisniku. Sve je to moguće uz uvet da je proxy prethodno registriran kod korisničke aplikacije(4).

2.3 Praktična strana

XML (Extensible Markup Language)

standardizovani jezik za opis podataka, koji omogućava programerima razvoj sopstvenih tipova podataka, u cilju identifikacije informacija u dokumentima, te njihovom stavljanju u odgovarajući kontekst.

SOAP (Simple Object Access Protocol)

predstavlja protokol koji opisuje način na koji jedna aplikacija treba da komunicira sa web servisom. SOAP omogućava korišćenje web servisa za finansijske transakcije na Internetu, kao što je recimo autorizacija kreditnih kartica.

UDDI (universal description, discovery and integration)

sistem sličan “yellow pages” sistemu, na kome su izlistani web servisi, a koji dopušta softverskim aplikacijam da “prelistaju” dostupne web servise i opisuje način kako da se “zakače” za njih.

WSDL (Web services description language)

jezik koji služi za definisanje pojedinačnih web servisa. WSDL jezikom opisuju se sve vrste aktivnosti koje može da odradi neki web servis

Praktično gledano, web servis predstavlja bilo koju softversku aplikaciju koja je dostupna preko mreže (Internet ili intranet) i koja koristi standardizovan XML sistem poruka. XML se koristi za formatiranje svih oblika komunikacije ka web servisima. Na primer, klijent poziva web servis slanjem poruke u XML formatu, zatim čeka na odgovarajući XML odgovor.

Pošto se sva komunikacija formatira uz pomoć XML standarda, web servisi nisu vezani ni za jedan operativni sistem ili razvojno okruženje - jednostavno, Java aplikacije mogu da "komuniciraju" sa Perl aplikacijama, dok recimo Windows programi mogu da razmenjuju podatke sa programiranim koji rade u Unix okruženju.

Web servisi sasvim prevazilaze nekompatibilnosti skupom softverskih standarda kao što su XML, SOAP (simple object access protocol), UDDI (universal description discovery and integration) i WSDL (Web services description language). Ovi standardi omogućuju definisanje, pakovanje, pristupanje i izvršavanje podataka i programa preko Interneta, bez potrebe za vođenjem računa o pojedinačnim implementiranim tehnologijama.

SOAP

SOAP je jednostavan protokol za razmenu informacija u decentraliziranom, distribuiranom okruženju. Baziran je na XML message protokolu koji je standardiziran od strane World Wide Web Consortium (W3C). Pomoću SOAP-a se može implementirati RPC stil komunikacije u kojem SOAP poruka sadrži pozivajuću funkciju i parametre koje prosljeđuje funkciji. Uobičajeno je da takvu poruku šalje korisnik, a poslužitelj vraća poruku s rezultatom pokrenute funkcije. SOAP takođe podržava "dokumentni stil" aplikacije u kojima je SOAP samo omotač oko XML dokumenta. Takve aplikacije su veoma fleksibilne i zato se danas koristi ta prednost prilikom kreiranja Webservisa.

SOAP dokument se sastoji od obavezne SOAP envelope, opcionalnog SOAP zaglavlja (header) i obaveznog SOAP tela (body).



SOAP Evenlope sadrži encondingStyle atribut koji definira korištena pravila serijalizacije u dotičnoj SOAP poruci. Taj se atribut može pojaviti u bilo kojem elementu i prenosi se na sve podređene elemente (child elements) koji ne sadrže vlastiti atribut.

SOAP Header je prvi podređeni element SOAP envelope elementa. Podređeni elementi header elementa (child elements) moraju biti identificirani sa namespace (obično se koristi URI) i lokalnim imenom. Svaki element može sadržavati atribut mustUnderstand kojim se primatelju naznačuje da je obvezno procesiranje zadržati (vrijednost atributa =1). SOAP poruka koja putuje do određene destinacije obično na svom putu prolazi kroz skup posrednika. Za tu svrhu postoji poseban atribut actor koji definira kojim je sve posrednicima namijenjena poruka.

SOAP Body element je podređeni element SOAP envelope elementa koji se koristi za prenos parametara metoda i (opcionalno) poruka o grešci.

SOAP je iznimno lako povezati s HTTP protokolom, koji je podržan od skoro svih operacijskih sistema. Infrastruktura za nadgledanje i sigurnost unutar HTTP-a je također već razvijena.

Prednosti Web servisa u odnosu na "klasične" Web aplikacije

Fleksibilniji su. Klasične Web aplikacije se mogu koristiti samo uz pomoć Web browser-a, pri čemu korisnik nema nikakvog uticaja na to kako će rezultat izgledati na njegovom ekranu, osim ako ne napravi program koji će da iz dobijenog HTML-a "izvlači" interesantne podatke i prikazuje ih na željeni način. Web servise mogu da koriste najrazličitiji programi: počev od drugih Web servisa, preko Web aplikacija, GUI programa ili čak tzv. command-line aplikacija. Recimo servis za prevod teksta koji smo pomenuli kao primer, može da koristi Altavista.com, program za obradu teksta kao što je MS Word, može se napraviti i programčić u Unix tradiciji koji bi radio u stilu "ttx -ef mytextfile" (prevod sadržaja fajla mytextfile sa engleskog na francuski), a mogao bi da bude korišćen i sa npr. mobilnih telefona ili ručnih računara koji imaju vezu sa internetom. Svi ovi različiti "klijenti" bi koristili jedan isti Web servis i autori tog servisa ne bi morali ni da znaju kakve sve različite aplikacije ga koriste.

Štedljiviji su po pitanju opterećenja mreže i resursa servera. Ako pogledamo primer sa Systran-ovom Web aplikacijom, videćemo da ona najpre mora da pošalje korisniku HTML sa formom za unos teksta, a posle prevoda šalje mu ne samo prevedeni tekst, već i HTML kod koji opisuje kako taj tekst treba da bude prikazan. Web servis bi primio jedan zahtev sa originalnim tekstom i poslao jedan odgovor sa prevedenim tekstom - nikakav dodatni HTML.

Lakši su za razvoj, testiranje i održavanje. Kod Web aplikacije je neophodno odraditi ne samo funkciju programa, već i dizajn i sve to dobro istestirati sa raznim browserima i platformama. Ko je probao da uskladi da Web aplikacija izgleda dobro na različitim browserima zna kakva je to muka (pitajte autora ovog sajta), a kad se napravi neka izmena, proces testiranja ide iz početka. Sa Web servisima svega toga jednostavno nema. Autor servisa brine o funkciji - prikaz je na klijentima.

Znači li to da će Web servisi zameniti Web aplikacije koje rade sa HTML-om?

Ne, osim u retkim slučajevima. Web servisi imaju svoj spektar primene koji je mnogo širi od onog koje pokrivaju Web aplikacije. Međutim, u mnogim slučajevima je još uvek bolje raditi sa HTML prikazom nego praviti specijalizovane programe koji će tumačiti podatke dobijene od servera. ES je primer Web aplikacije koja treba da ostane u ovom obliku, mada nije nemoguće dodati i Web servis.

2.4 Primeri upotrebe web servisa

Kao jedan od najjednostavnijih primera upotrebe web servisa možemo uzeti aukcijski endžin poznate Interneta aukcijske kuće eBay [www.ebay.com]. Zainteresovane firme i pojedinci mogu se jednostavno pretplatiti na korišćenje eBay-evih web servisa, tako da ne moraju investirati u razvoj sopstvenih. Sve što je potrebno je da u svoje web aplikacije implementiraju nekoliko redova programskog koda kako bi inkorporirali eksterne web servise.

Na taj način je praktično omogućeno da druge kompanije (koje u osnovi nisu aukcijske kuće) imaju funkcionalnost jedne aukcijske kuće, te mogu da organizuju privatne instant-aukcije na spostvenim web sajtovima. Dodatno ovaj web servis može biti potpomognut servisima drugih kompanija koje se bave procesiranjem plaćanja, isporuke i logistike, tako da zainteresovane firme (pojedinci) mogu da kreiraju aukcijske portale bez sopstvenih resursa, a u saradnji sa brojim provajderima različitih web servisa.

Još jedan vrlo plastičan primer upotrebe web servisa je kompanija CUNA Mutual [www.cuna.org], koja se bavi pružanjem finansijskih usluga kreditnim ustanovama. Ova kompanija koristi web servise za transfer podataka iz back-end aplikacija za upravljanje finansijskim servisima na web stranice. Većina ovih softverskih aplikacija je urađena pre više godina, kada se još nije naslućivao veliki potencijal Interneta. U kompaniji su procenili da nije isplatljivo ponovo razvijati iste aplikacije koje bi nudile dodatne dinamičke sadržaje. Umesto toga, odlučili su da kreiraju web servise koji automatski formatiraju podatke iz pozadinskih aplikacija tako da vrlo jednostavno mogu (preko content-management sistema) da se prikažu na web stranicama.

2.5 Potencijali u domaćim uslovima

Kao što je gore navedeno, web servisi, njihov razvoj i upotreba, mogu biti najzanimljiviji finansijskom sektoru, tj. domaćim bankama, štedionicama, kreditorima, brokerskim kućama i sličnim ustanovama. Upotreba ove nove tehnologije još više dobija na značaju prelaskom platnog prometa u poslovne banke. Kreiranjem odgovarajućih web servisa i usaglašavanjem komunikacije između poslovnih partnera shodno XML standardima povećava se efikasnost i operativnost, smanjuju se troškovi i ojačava se veza između poslovnih partnera i klijenata.

Banalni primer opravdanosti uvođenja web servisa odnosi se na kreiranje i distribuciju izveštaja o stanju na računima klijenata, popularnim "izvodima". Trenutno, domaće banke ove izveštaje nude u raznoraznim oblicima, a najčešće u štampanom obliku ili u obliku tekstualnih datoteka različitih formata. Ovako formirane informacije se vrlo teško unose u informacione sisteme klijenata, što uvećava troškove i kompleksnost. Rešenje bi bilo formiranje web servisa koji bi opsluživao klijente podacima koji su standardizovanog formata. Ovim bi takve poruke mogle, bez dodatnih transformacija i konverzija, da ulaze u poslovne informacione sisteme klijenata.

U perspektivi, poslovni partneri između sebe i u komunikaciji sa bankama, bi trebali da razviju sopstvene web servise kako bi standardizovali komunikaciju, te povećali efikasnost i konkurentnost. Sa web servisima, poslovanje se premešta sa mreže koja je oslonjena na ljudski faktor, na mrežu koja se oslanja na aplikacije (aplikativno-orijentisane mreže gde međusobno komuniciraju aplikacije). Ovo ne znači da se iz priče izbacuje ljudski faktor, već da se samo omogućava direktna konverzacija između aplikacija, kao što to sada imamo u pojednostavljenom obliku između web browser-a i web servera (zahtev-odgovor).

3. Razvoj web servisa korišćenjem .Net platforme

3.1 Uvod

Nakon više godina ozbiljnog razvoja i pompeznih najava, krajem 2006. –te godine, Microsoft je napokon izbacio na tržište svoju novu platformu za razvoj Windows aplikacija. Jezgro te platforme počiva na .NET Framework-u 3.0. Bitno je napomenuti da .NET Framework 3.0 nije prost naslednik .Net Framework-a 2.0 već se pre može posmatrati kao njegovo proširenje. Samo po sebi postavlja se pitanje u čemu se ogleda to proširenje. Odgovor se krije u činjenici da .NET Framework 3.0 u sebi sadrži nekoliko potpuno novih biblioteka, koje do sada nisu postojale u prethodnim verzijama .NET framework-a. Korišćenjem ovih biblioteka moguće je efikasno razviti potpuno novu generaciju aplikacija na Windows platformi.

Najvažnije biblioteke koje ulaze u sastav .NET Framework 3.0 su:

- ✓ Windows Presentation Foundation (bivši radni naziv je Avalon),
- ✓ Windows Communication Foundation (bivši radni naziv je Indigo),
- ✓ Windows Workflow Foundation
- ✓ Windows Card Space

Ove biblioteke predstavljaju srž ozbiljne razvojne platforme na kojoj Microsoft zasniva razvoj aplikacija za Windows operativne sisteme u narednom periodu. Operativni sistemi na kojima aplikacije razvijene uz pomoć .NET Framework-a 3.0 mogu da rade su:

- ✓ WindowsXP SP 2,
- ✓ Windows2003 Server,
- ✓ Windows Vista

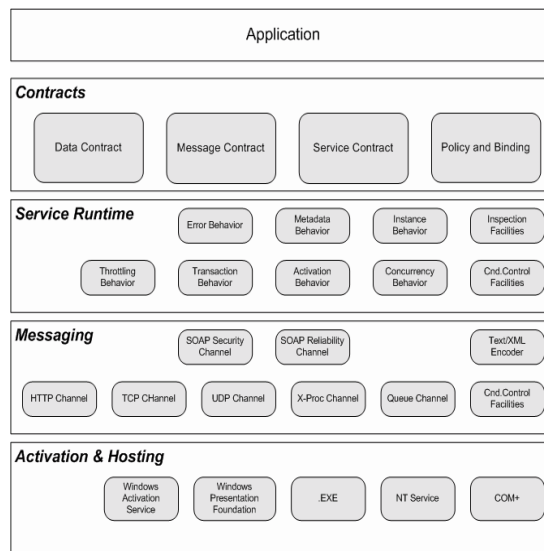
U daljem tekstu ćemo se detaljnije pozabaviti arhitekturom WCF-a, razvojem WCF aplikacija, kao i deployment-om WCF aplikacija.

3.2 Osnove windows communicatio foundation-a

Windows Communications Foundation (WCF) je nov Microsoft-ov API za razmenu poruka. WCF omogućava programerima dizajniranje i razvoj sigurnih i pouzdanih Web servisa u heterogenim okruženjima. Takođe WCF omogućava jedinstveni API za razmenu sigurnih poruka između različitih aplikacija, bez obzira da li se komunikacija vrši kao inter-process komunikacija (IPC), preko LAN-a ili preko Interneta. Napredna podrška za Web servise ugrađena u WCF obezbeđuje razmenu sigurnih transakcionih poruka. WCF je servisno orjentisan programski model, izgrađen nad .NET Framework-om 2.0 i značajno pojednostavljuje izgradju konektovanih sistema. WCF je direktan naslednik nekoliko postojećih Microsoft-ovih proizvoda:

- ✓ ASP.NET web servisa (ASMX) i Microsoft Web Services Enhancements for Microsoft .NET (WSE 3.0)
- ✓ .NET Remoting
- ✓ Enterprise Services
- ✓ System.Messaging

Funkcionalnost napred nevedenih proizvoda WCF objedinjuje u jedinstvenu biblioteku.



Slika 1. Arhitektura WCF-a

WCF omogućava čitav spektar scenarija pri razmeni poruka kao što su :

- ✓ One-way i duplex razmena poruka
- ✓ Sinhroni i asinhroni pozivi udaljenih procedura
- ✓ Callback-ovi
- ✓ Sesije
- ✓ Multi-contract servisi
- ✓ Transport-based i message-based security, reliability, and ordered delivery
- ✓ Queued messaging
- ✓ Podrška za transakcije
- ✓ Composability

Osnovne prednosti koje WCF donosi u odnosu na dosadašnji način razvoj servisa su sledeće:

- ✓ Jedinstven programski model koji će obezbediti veću produktivnost pri razvoju aplikacija i jednostavnije održavanje servisa.
- ✓ Razvoj i modifikacija servisa nezavisno od protokola na kome se poruke prenose.
- ✓ Bolje performanse (WCF servisi su 5 do 10 %
- ✓ brži od ASMX servisa)

- ✓ WCF je Microsoft-ova implementacija svih WS-* standarda.

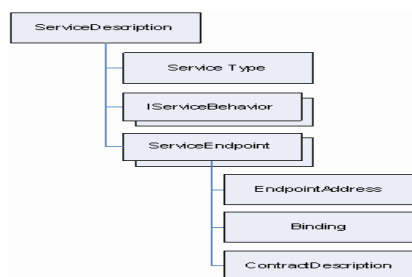
WCF servisi mogu biti ugrađeni u širok spektar aplikacija. Jedna od najzanimljivijih karakteristika je da Web servisi mogu na jednostavan način biti hostovani u okviru konzolskih aplikacija, klasičnih desktop aplikacija, windows servisa ili web aplikacija.

3.3 Osnove WCF klase i pojmovi

WCF klase se najvećim delom nalaze u okviru System.ServiceModel namespace-a ili namespace-ova koji njemu pripadaju.

ServiceDescription klasa opisuje WCF servis. Ona sadrži tri osnovne komponente :

- tip servisa ,
- kolekciju behaviors-a servisa i
- kolekciju ServiceEndpoint objekata.



Slika 2. Osnovni elementi ServiceDescription klase

Tip servisa je tip koji ServiceDescription opisuje. Sve dolazeće poruke se dispečuju u instance tog tipa radi procesiranja. Servis behaviour je tip koji implementira IServiceBehavior interfejs definiše runtime funkcionalnost servisa. Endpointi su osnovne jedinice komunikacije u WCF-u. Oni su predstavljeni klasom ServiceEndpoint. Svaki endpoint je sačinjen od tri glavna elementa:

- adresa
- binding
- contract

Adresa je predstavljena instancom klase EndpointAddress. EndpointAddress klasa sadrži u sebi URI, identity (koji se koristi iz razloga bezbednosti), i kolekcije opcionih zaglavlja. Opciona zaglavlja se koriste da sačuvaju dodatne adresne informacije.

Binding definiše transportni protokol, encoding, određuje da li endpoint održava stanje (sesije) i da li učestvuje u transakcionim procesima. HttpTransportBindingElement određuje transportni protokol koji će endpoint koristiti, ReliableSessionBindingElement određuje da li će endpoint

podržavati sesije, SecurityBindingElement određuje security aspekt endpointa. WCF isporučuje sa skupom predefinisanih bindinga koji uključuju najveći broj potrebne funkcionalnosti. Takođe je moguće napraviti i sopstvene bindinge.

Contract je kolekcija operacija koje određuju kako endpoint komunicira s spoljnim svetom. ContractDescription klasa se koristi da opiše WCF endpoint contract-e i njegove operacije. U okviru ContractDescription klase ,svaka operacijama odgovarajuću OperationDescription klasu koja opisuje aspekt operacije , na primer da li je operacija jednosmerna ili request/reply. Svaki OperationDescription element takođe opisuje poruke koje sačinjavaju operaciju koristeći kolekciju MessageDescription objekata.

3.4 Osnovne osobine endpointa

Servis može da ima više endpoint-a (objavljenih contract- a na određenoj adresi) u isto vreme. Contract može da podrži više binding-a i binding može da podrži više contract-a. Contract može da ima specifične zahteve vezane za binding kao što su zahtevi vezani za bezbednost, transakcije itd... Ako binding ne zadovoljava uslove contrac-a on neće biti startovan.

3.5 Donošenje odluke o korišćenju binding-a pri razvoju servisa

Pri razvoju WCF aplikacija jedna od glavnih odluka koju je potrebno doneti je, koje ćemo binding-e koristiti. Ta odluka zavisi od nekoliko osnovnih karakteristika servisa koje želimo da razvijemo. Prva karakteristika je kakav nivo interoperabilnosti servisa želimo, t.j. da li želimo da se servis koristi i od strane klijenata razvijenih i u drugim tehnologijama ili samo od strane klijenata razvijenih u .Net tehnologiji. Druga bitna karakteristika je, kakav tip security-ja želimo da koristimo, da li na nivou transportnog protokola, da li na nivou poruke ili mešoviti mod. Treća karakteristika je da li želimo da naš servis podržava sesije ili ne. Četvrta karakteristika je da li želimo da naš servis podržava transakcije ili ne. Peta karakteristika je da li želimo da koristimo duplex poruke ili ne. Na donjoj slici je moguće videti koji sve tipovi binding-a postoje u okviru WCF-a i koje od gore navedenih pet karakteristika podržavaju.

Binding	Interoperability	Security (Default)	Session (Default)	Transactions	Duplex
BasicHttpBinding	Basic Profile 1.1	(None), Transport, Message	None, (None)	(None)	n/a
WSHttpBinding	WS	Transport, (Message), Mixed	(None), Transport, Reliable Session	(None), Yes	n/a
WSDualHttpBinding	WS	(Message)	(Reliable Session)	(None), Yes	Yes
WSFederationHttpBinding	WS-Federation	(Message)	(None), Reliable Session	(None), Yes	No
NetTcpBinding	.NET	(Transport), Message	Reliable Session, (Transport)	(None), Yes	Yes
NetNamedPipeBinding	.NET	(Transport)	None, (Transport)	(None), Yes	Yes
NetMsmqBinding	.NET	Message, (Transport), Both	(None)	(None), Yes	No
NetPeerTcpBinding	Peer	(Transport)	(None)	(None)	Yes
MsmqIntegrationBinding	MSMQ	(Transport)	(None)	(None), Yes	n/a

Slika 3. Osnovni karakteristike binding-a

Iz gornje tabele vidi se da ukoliko želimo da razvijamo interoperabilne servise možemo da koristimo: BasicHttpBinding, WsHttpBinding i WsDualHttpBinding.

Ukoliko želimo da razvijamo servise koji će se koristiti iz WCF klijenata onda ćemo koristiti: NetTcpBinding, NetNamedPipeBinding, NetMsmqBinding, NetPeerTcpBinding.

3.6 Razvoj aplikacija u .NET FRAMEWORKU 3.0

Postavljanje razvojnog okruženja

Da bi ste na mašini na kojoj je instaliran Windows XP mogli nesmetano da počnete razvoj .Net framework 3.0 aplikacija najpre je potrebno instalirati sledeće proizvode:

- ✓ Windows XP , Service pack 2
- ✓ NET framework 2.0
- ✓ Net Framework 2.0 SDK
- ✓ Visual Studio 2005
- ✓ Visual Studio 2005 SP1

Potom je potrebno instalirati sledeće proizvode:

- ✓ Microsoft .NET Framework 3.0 Runtime
- ✓ Componenets
- ✓ Microsoft .NET Framework 3.0 SDK
- ✓ Visual Studio 2005 Extensions for .NET framework 3.0 (WCF & WPF) – trenutno je dostupan Novembar 2006 CTP
- ✓ Visual Studio 2005 Extensions for .NET framework 3.0 (Windows Workflow Foundation)

Nakon uspešne instalacije pomenutih proizvoda moguće je započeti razvoj NET 3.0 aplikacija. Napominjemo da je razvojno okruženje moguće postaviti i na mašinama koje su pokrenute pod Windows Vista operativnim sistemom. Proces postavljanja razvojnog okruženja izgleda slično kao i proces postavljanja razvojnog okruženja na Windows XP-u, ali se u ovom tekstu nećemo detaljnije baviti ovim procesom. Za sve detalje možete pogledati [1] i [2].

Pokretanje .NET 3.0 projekata iz Visual Studija 2005

Nakon uspešne instalacije gore pomenutih proizvoda, u Visual Studiju 2005, će se u New Project dijalogu, sa leve strane pojaviti odeljci: NET framework 3.0 i Workflow. Klikom na NET framework 3.0 odeljak sa desne strane se prikazuju različiti tipovi projekata:

- Windows Application (WPF)
- XAML Browser Application
- Custom Control Library
- WCF Service Library

Nakon klika na bilo koji od ovih ikonica u VisualStudiju će se otvoriti novi projekat odgovarajućeg tipa.

Da bi ste započeli razvoj WCF servisa potrebno je kliknuti na ikonicu sa naslovom WCF Service Library.

U ovom članku nećemo se baviti konkretnim razvojem WCF servisa. Mnoštvo detalja i korisnih primera vezanih za razvoj WCF servisa možete naći u [3].

3.7 Napomene

Iako je .NET Framework 3.0 još uvek relativno nov proizvod, mišljenje je, da se aplikacije razvijene u njemu ponašaju veoma stabilno. Pored toga .NET Framework 3.0 pruža i čitav spektar novih mogućnosti koje značajno mogu da unaprede aplikacije koje razvijate. Takođe, nove ekstenzije za Visual Studio 2005 omogućavaju da se uz pomoć tog integrisanog razvojnog okruženja na efikasan način razvijaju NET 3.0 aplikacije. Pri tome će razvoj ovih aplikacija biti u velikoj meri konzistentan sa dosadašnjim načinom razvoja bilo Windows aplikacija bilo ASMX web servisa, što omogućava programerima da se relativno bezbolno prabace na novu tehnologiju i pri tome iskoriste sve njene prednosti.

Stoga je preporuka da ukoliko do sada niste probali razvoj sa ovom tehnologijom, pravi je trenutak da počnete.

U ovom delu rada smo objasnili osnovnu namenu .NET 3.0 framework-a i naveli glavne biblioteke koje su u njega uključene. Potom smo se fokusirali na WCF i detaljnije objasnili njegovu strukturu i osnovne pojmove bitne za njegovo razumevanje. Potom smo objasnili na koji način je moguće postaviti razvojno okruženje za razvoj .NET. Na kraju smo se dotakli i načina deployment-a WCF servisa i dali neke dodatne napomene. Microsoft sa pravom polaže dosta nade u .NET framework 3.0. Uz pomoć biblioteka koje su u njega uključene značajno će se ubrzati i olakšati razvoj i održavanje kompleksnih i zahtevnih softverskih sistema. Takođe, novi Microsoft-ovi operativni sistemi (npr Windows Vista), svoju funkcionalnost uveliko zasnivaju na .NET 3.0 komponentama. Za uspeh jedne tehnologije, veoma je bitna i činjenica ko je njen proizvođač i sa koliko je upornosti i vere promoviše, usavršava i unapređuje. Sudeći po Microsoft-ovim dosadašnjim aktivnostima, NET framework 3.0 tehnologija je nešto što će oni uporno promovisati i neprestano usavršavati. Stoga je mišljenje, da će .NET framework 3.0 komponente, u veoma bliskoj budućnosti, imati presudan uticaj na razvoj novih, kvalitetnih, softverskih rešenja i da će veoma brzo naći svoju primenu u širokom spektru komercijalnih poslovnih aplikacija. U takvim poslovnim aplikacijama WCF servisi će zasigurno imati jednu od najznačajnijih uloga.

4. Osnovni servisi

Osnovni servisi su prisutni na svakom računaru koji je povezan sa Internetom. Da bi se koristili potrebni su : korisničko ime na nekom od servera i lozinka za pristup.

E-mail (elektronska pošta) – ovaj servis obezbeđuje slanje poruke drugom korisniku. Za razliku od drugih servisa e-mail se može koristiti čak i kad ne postoji stalna veza sa Internetom. Elektronska pošta je najstariji servis ali i dalje jedan od najpopularnijih. Poruka koja se šalje stiže za vrlo kratko vrijeme, a primljena poruka može se dalje obrađivati. Osim teksta mogu da se prenose i slike, formulari ili bilo kakav dokument.

Karakteristike:

- a) istovremeno slanje na veći broj adresa,
- b) provjera autentičnosti poruke,
- c) korištenje nadimka umjesto adrese,
- d) slanje više kopija iste poruke,
- e) automatsko kreiranje nove poruke,
- f) automatsko odgovaranje na primljenu poruku,
- g) nedostatak automatske provjere da li je poruka isporučena ili pročitana,
- h) postojanje mehanizma fascikli radi lakše organizacije pošte,
- i) izdvajanje poruka,

4.1 Javni servisi

Grupe ovih servisa instalirane su na značajnim serverima u mreži da bi se korisnicima obezbedio jednostavan pristup podacima bez dodatne provjere identiteta. Kao što su na primer:

- Mailing list – predstavlja proširenje elektronske pošte za komunikacione grupe korisnika koje imaju slična interesovanja. Poruka poslana na listu stiže svim korisnicima koji su na nju pretplaćeni.
- Anonimus FTP – to su u suštini javna skladišta datoteka do kojih dolazimo standardnim programom ftp. Datoteke su dostupne svim korisnicima. Na nekim od servera postoji poseban korisnik pod imenom "anonimus" koji nema lozinku, a u čijem katalogu se nalaze datoteke koje svi mogu da čitaju ali ne mogu da upisuju podatke.

Primer: <ftp.microsoft.com> <ftp.mastercard.com>

- Usnet News (User Network News) – predstavljaju javne diskusione grupe ili konferencije (BBS). Poruke su dostupne svim korisnicima mreže. Sve poruke su organizovane po diskusionim grupama i lancima diskusije. Kad se čitaju poruke prvo se bira diskusiona grupa a zatim lanac diskusije i konkretne poruke. Diskusionih grupa u svetu ima preko 5000, ali nisu sve dostupne na svim serverima , njihov broj zavisi od zahtjeva korisnika, kapaciteta diskova i brzine veze. Usled velikog broja diskusionih grupa i nemogućnosti snalaženja uveden je koncept pretplate (subscribe). Svaki korisnik bira one grupe koje ga

interesuju i pri pristupu se pojavljuju samo te grupe. Postoji veliki broj programa čitača za pristup servisu Usenet News od kojih su najpoznatiji: rn, trn, tin na Unix-y, ano-news na VMS-u, a pod Windows-om: Netscape, Internet Explorer i drugi koji imaju mogućnost rada sa raznim servisima Interneta (e-mail, Usnet News i dr.).

- Gopher – u suštini predstavlja sistem menija kroz koje se pristupa tekstualnim datotekama kao i raznim serverima u mreži. Kada se na nekom računar pokrene Gopher on kontaktira odgovarajuće servere sa kojih dobiva menije i datoteke za prikazivanje. Datoteke su najčešće tekstualne, ali ako se radi u prozorskom sistemu mogu da se vide slike i drugi multimedijalni dokumenti (xgopher). Gopher-y se može pristupiti ili preko posebnih programa (ako je operativni sistem UNIX otkuca se gopher u komandnoj liniji) ili preko bilo kog WWW čitača (Netscape Navigator, Internet Explorer, na početku se umjesto "http://" otkuca "gopher://").
- WWW - World Wide Web (spider web -paukova mreža) je najpopularniji servis na Internetu kojim se dokument na mreži povezuje u obliku hiper-veza. Osnova dokumenta je tekst u kome se mogu nalaziti drugi multimedijalni sadržaji. Protokol kojim WWW klijent i server komuniciraju naziva se HTTP (Hiper Text Transfer Protocol). Preko WWW-a može se pristupiti i drugim serverima na Internetu kao što su: drugi WWW server, Anonimus FTP, Gopher, News, Telnet, i mnogi drugi.
- JAVA - predstavlja proširenje WWW-a koje omogućava da se preko mreže prenose programi koji se izvršavaju na našem računaru. JAVA je programski jezik sličan C++. Pomoću ovog programskog jezika može da se crta, kontrolišu ekranske forme, ispisuje tekst i slike te slične stvari. Sve se to obavlja na lokalnom računaru, što je brže nego kada se radi preko mreže. Smatra se da će jezik JAVA biti osnova razvoja tehnologije za elektronsku trgovinu.
- IRC (Internet Realy Chat) - ovaj servis omogućava interaktivni razgovor velikog broja korisnika. Izmišljen je u Finskoj i postao jedan od najpopularnijih servisa. Celi sistem je organizovan po kanalima koji se dele na : javne, privatne, nevidljive i tajne.
- Mail gateways - servisi su namijenjeni pristupu javnim servisima elektronskom poštom. Omogućava se pristup i onim korisnicima koji nemaju stalnu vezu sa Internetom.

4.2 Servisi za pretraživanje

Ova grupa servisa omogućava pronalaženje odgovarajućih dokumenata na javnim servisima. Bez njih bi kretanje Internetom bilo znatno sporije. U ovu grupu servisa spadaju:

- ◆ Archie – služi za pretraživanje Anonimus Ftp-a. Do njega se može doći na više načina:
 - a) preko specijalnih klijenata,
 - b) preko servisa telnet (telnet archie.rytgers.mcgill.ca),
 - c) preko elektronske pošte, ako nisu potrebni podatci trenutno (archie@archie.mcgill.co),
 - d) preko WWW-a (<http://www-ns.rutgers.edu/htbin/archi>)

- ◆ Veronica – ovaj program namenjen je pretraživanju Gopher-a. Mana ovog servisa je dugo čekanje na odgovor usled prelistavanja gopher-a u celom svetu. Zbog ovog nedostatka stvoren je poseban servis Jughead.

- ◆ Wais (Wide Area Information Service)- je nastao zbog potrebe pretraživanja tekstova. Ovaj servis proverava unapred pripremljenu bazu podataka, pronalazi dokumente koji se poklapaju sa zadanim upitom i ispisuje njihovu listu zajedno sa svojom ocenom o nivou poklapanja. Ako ima previše pronađenih dokumenata, upit se zadaje preciznije. Program Wais je postao javno dostupan, a formirana je i organizacija CNIDR (Clearinghouse for Networked Information Diskovery Retrieval) koja se bavi održavanjem liste. Postoje i drugi sistemi za pretraživanje mreže kao što su: Lycos, Yahoo, Alta Vista, Magelan i drugi.

- ◆ Na WWW-u postoji dosta servisa koji vrše pretraživanje, jedan od najpoznatijih i najstarijih je Yahoo čija je adresa:

<http://www.yahoo.com>

Yahoo je organizovan po principu sadržaja i stranice su razvrstane po tematici.

- ◆ Alta Vista je drugi tip Internet pretraživača čija je adresa:

<http://www.altavista.com>

Ovaj servis podržavaju tri brza Digital Alpha servera sa specijalnim softverom koji 24 časa dnevno pretražuje i indeksira sav sadržaj WWW-a i Usenet konferencija. Veoma je lak i brz za korišćenje. Postoji još Internet pretraživača (Lycos, WebCrawler) koji nisu tako popularni kao Yahoo i Alta Vista, ali nude neke specifične usluge.

- ◆ Google – kao jedan od najpopularnijih Internet servisa koji se danas primenjuju u svakodnevnoj praksi. Njegova adresa je:

<http://www.google.com>

Google.com je nastao kao akademski pretraživač. Njegov inicijalni sistem koristi više paukova, obično tri istovremeno. Svaki pauk može da održi 300 veza sa Web stranicama otvorenim u isto vreme. Pri svojim maksimalnim performansama, koristeći četiri pauka,

sistem može da odgmiže preko 100 stranica po sekundi, stvarajući oko 600 kilobajta podataka svake sekunde. Održavajući sve brzim značilo je napraviti sistem koji bi hranio pauke neophodnim informacijama.

Prvobitni sistem Google-a sadržao je server posvećen za dostavljanje URL-a paucima. Google je imao svoj sopstveni DNS (Domain Name) Server (koji prevodi ime servera u adresu), iz razloga što bi se u suprotnom oslanjao na Internet servis provajdera za DNS. Rezultat je svođenje odlaganja na minimum, tj. veća brzina.

Kada Google-ov pauk pretražuje HTML stranice, on beleži dve stvari:

- Reči na stranicama
- Lokaciju reči

Reči koje se pojavljuju u naslovu, podnaslovu, meta tagovima i drugim važnim pozicijama su zabeležene za specijalno razmatranje u toku sedeće korisničke pretrage. Google-ov pauk pravi index svake značajne reči na stranici, ostavljajući članove “a”, “an” i “the”. Drugi pauci koriste drugačije pristupe.

Drugačije pristupi su pokušaj da pauk radi brže, dozvoli korisniku da pretražuje efikasnije ili oboje. Na primer neki paukovi će čuvati trag reči u naslovu, podnaslovu i linkovima, zajedno sa sto najčešće korišćenih reči na stranici i svaku reč u prvih dvadeset redova teksta. Tvrdi se da Lycos koristi ovakav pristup prilikom pretrage Web-a paucima.

- ♦ Netfind - služi za pronalaženje e-mail adrese korisnika na osnovu imena i prezimena. Pretraživanje se vrši komandom finger.

4.3 Sigurnosni servisi

Sigurnosni servisi su neophodni kada želimo da zaštitimo podatke koji se prenose preko Interneta. Nabrojaćemo neke od takvih servisa:

- PGP – koji tehnikom javnog i tajnog ključa šifrira elektronsku poštu.
- SSH – služi da šifrira interaktivnu komunikaciju.
- KERBEROS – koji ima ulogu da obezbeđuje vezu sa računarom koji proverava identitet servera i korisnika.

4.4 Sistemski servisi

Ovi servisi namenjeni su administratorima servera i mreža radi provere da li su računari priključeni na mrežu i da li postoje problemi sa vezama. Spomenućemo neke od tih servisa:

- ❖ Ping – on nam služi za proveru postojanja i kvaliteta veze među računarima u mreži.
- ❖ Nfs – obezbeđuje podelu prostora na mrežnim diskovima. Pa u ovom slučaju veći broj računara može da pristupi istim delovima diska na serveru.

- ❖ Traceroute – njegova uloga je da proverava vezu i ispituje putanju kojom paket putuje od servera do servera.
- ❖ Nstate – ovaj servis trenutno prikazuje aktivne mreže na serveru.
- ❖ X-Windows – je distribuiran prozorski sistem.

Kao što smo ranije spomenuli, Internet je odlično mesto da se dobije informacija preko kontakata sa drugim korisnicima ili nalaženjem informacija iz različitih izvora.

4.5 Elektronska pošta

Elektronska pošta (Elektronic Mail, E-Mail) je nastao 1971. godine, a koncipirao ga je američki inženjer Ray Tomlison. Glavne prednosti elektronske pošte se ogledaju u jednostavnosti upotrebe u svakodnevnoj komunikaciji kako poslovnih, tako i ostalih korisnika. Elektronska pošta ne zahteva istovremeno prisustvo oba komunikatora. Poruke putuju daleko brže nego što je to slučaj kod tradicionalne pošte, dok je nivo pouzdanosti isporuke na odredište daleko veći. Velika prednost je i u njenoj jeftinoći u odnosu na ostale oblike pismene komunikacije (fax, pisma i sl.).

Putem e-maila komunikacija se može uspostaviti između bilo koja dva subjekta koji su prijavljeni kod provajdera Internet usluga i kojima je dodijeljena odgovarajuća adresa. Adresa se sastoji od:

Ime-korisnika@Internet-adresa

marko_markovic@yahoo.com

Ime korisnika korisnik određuje proizvoljno (po vlastitoj želji), a Internet adresu dodeljuje provajder. Komunikacija elektronskom poštom se odvija na bazi klijent/server modela.

Danas su najčešće u upotrebi standardni programi za rad sa elektronskom poštom poput Microsoft Outlook Express, a slabije se koriste Netscape Messenger, Eudora i Pegasus. Osim ovih postoji i niz drugih manje popularnih programa iste namene.

Najvažniji problemi sa elektronskom poštom vezani su za:

- 1) verifikaciju dostave
- 2) autorizaciju, i
- 3) privatnost.

X.400 OSI e-mail standard za razmenu obezbeđuje mehanizam za proveru da li je poruka primljena ili pročitana. Niko ne može garantovati da poruku stvarno šalje osoba čije je ime navedeno, a takođe i tekst može biti promenjen. Elektronska pošta može biti prekinuta ili pročitana bilo gde duž svog puta od jedne osobe do druge. Zatim je tu problem što primalac može da prosledi vašu poruku drugim ljudima bez vašeg znanja itd.

Standardi za elektronsku poštu:

Internet e-mail koristi "store-and-forward" model za razmenu poruka. Jedan ili više računara može da čuva i prosljeđuje poruke dok ne stignu do destinacijskog računara.

- SMTP (Simple Mail Transport Protocol)- (protokol za slanje izlaznih poruka) je dodatak standardnom formatu za tekstualne poruke. Ovi formati su limitirani , tako da mogu da sadrže samo jednu poruku, sastavljenu od ASCII simbola, a ta poruka ne sme da premaši izvesnu dužinu.

- MIME (Multipurpose Internet Mail Extension)- odobren je 1992 godine i dostigao je široku primenu. MIME poruke mogu sadržavati višestruke objekte, uključujući podršku za sliku, audio, video i multimedijalne poruke.

- POP (Post Office Protocol)- u suštini predstavlja protokol za primanje ulaznih poruka.

4.6 WORLD WIDE WEB

Poznat je još i kao: WWW, Web, W3. Veoma jako rasprostranjen i veoma popularan Internet servis, koji se po svom sadržaju može smatrati gigantskom bibliotekom multimedijalnih dokumenata.

U suštini WWW je mrežni servis ili infrastruktura za organizovanje informacija, dostupnih korisnicima Interneta u vidu multimedijalnih dokumenata ili Web stranica. WWW se može smatrati multimedijalnim sistemom bez ikakvih ograničenja. Kolekcija multimedijalnih Web stranica naziva se Web sajt (Web Site). Početna stranica Web sajta je naslovna stranica koja pruža informacije o sadržaju Web sajta.

Da bismo objasnili način funkcionisanja WWW-a, moramo objasniti neke osnovne pojmove:

- ⇒ HyperText Transport Protocol (HTTP)- je osnovni protokol koji se na Web-u koristi za transfer informacija. HTTP se koristi za komunikaciju između Web servera i Browser-a te on upravlja celokupnom interakcijom između klijenta i servera. Korisnik kretanjem po Web-u , ne zna gde se nalaze računari koji šalju tražene podatke, te informacije obrađuju Browser-i.
- ⇒ HyperText Markup Language (HTML)- omogućava standardni prikaz stranice, tj. da se Web stranice ne bi razlikovale u zavisnosti sa kojeg se Browser-a pristupa, potreban je standardni opis dokumenta, a to je ono što omogućava HTML.
- ⇒ Uniform Resource Locators (URL)- predstavlja tačno specifikiranu adresu sa koje se informacije primaju ili na koju se šalju. URL predstavlja lokaciju resursa kojima korisnik želi pristupiti. Adresa URL-a se razlikuje od zavisnosti od tipa resursa kome se pristupa, a koji se označava na početku URL adrese. Nabrojaćemo nekoliko URL adresa različitih tipova:

Telnet telnet://...

Gopher gopher://...

World-Wide-Web http://www...

E-mail miki@eunet.rs

USENET News news:alt.internet.services

Internet Browser-i

Internet Browser predstavlja program koji je potrebno pokrenuti pre nego što korisnik pristupi Internetu. Ovaj program omogućava korisniku prikazivanje Web stranica kao i navigaciju po mreži, tj. među Web stranicama. Zatim još omogućavaju da dobivene informacije sačuvamo za kasniju upotrebu, da pošaljemo određene informacije, kao i razne druge opcije. Prvi browser koji je razvijen bio je NCSA MOSAIC. Browser-e možemo razlikovati i po okruženju u kome rade: Browser-i iz Dos okruženja, i Browser-i iz WINDOWS okruženja. Danas prednost imaju Browser-i iz Windows okruženja, pre svega "Netscape Navigator" i "Internet Explorer".

Servisi (uopšteno)

Najpopularniji korisnički servisi koje Internet tehnologija omogućuje su:

- Elektronska pošta (e-mail) - mogućnost brze korespondencije sa bilo kime u svetu.
- Pretraživanje mreže i grafička prezentacija informacija (browsing) - mogućnost dostupa velikoj količini informacija, pretraživanje po kategorijama dobara, ključnim rečima i sl. te multimedijaska prezentacija podataka korisniku.
- Prisustvo na mreži (Internet presence) - prezentacija vlastitih proizvoda i ideja, uključivanje informacija o vlastitim proizvodima u svetske baze podataka čime oni postaju dostupni svakome. Poseban oblik prisustva na mreži je i prodaja putem Interneta te ostali oblici novčarskih transakcija.
- "Okrugli stolovi", organiziranje korisnika Interneta zainteresiranih za pojedina područja u organizirane skupine u kojima se razmenjuju vredne informacije, novosti, komentari i pogledi na područje interesa.
- Dostup datotekama i mogućnost distribucije istih Internetom (mogućnost da distribuirate nove inačice svojih programa mrežom, da dobavite i isprobate demo verzije programa i sl.) .

E-mail

Elektronska pošta osigurava brzu i efikasnu korespondenciju sa svetom. Poslana poruka u trenu se pojavljuje na ekranu onoga kome je namenjena, on Vam istoga trena može odgovoriti ili jednako brzo proslediti Vašu poruku na jednu ili više drugih adresa. Pošta može uključivati i elemente multimedije, ton, sliku, pa i pokretnu sliku ako baš želite. E-mail program na Vašem računaru brinut će o tome da vas upozori da je pristigla pošta, da neko od ranije pristiglih elektronskih pisama niste pročitali, da li je pismo koje ste poslali pristiglo na pravu adresu. Vaš će elektronski pismonoša naravno voditi računa i o arhiviranju važnije pošte. I sve to uz apsolutnu sigurnost - Vaše poruke mogu putovati telefonskim i satelitskim linijama po svetu u šifriranom obliku tako da nitko osim onoga kome su namenjene ne može zaviriti u tako formiranu elektronsku omotnicu.

5. Registracija web servisa UDDi Registri

5.1 Pronalaženje web servisa na Internetu

Da bi pomogli programerima u dizajniranju i upotrebi Web servisa, Qwest Communications International Inc. je razvio Universal Description Discovery and Integration (UDDI) servis u Microsoft® Windows Server™ u okviru njihove organizacije. UDDI je standard za objavljivanje i pretraživanje postojećih servisa na mreži. Ovaj rad daje informacije kako bi UDDI trebao da se uklopi u arhitekturu Web servisa, koje su njegove mogućnosti, a koje prednosti, koji su uslovi potrebni da bi se mogao koristiti i na koji način se upotrebljava.

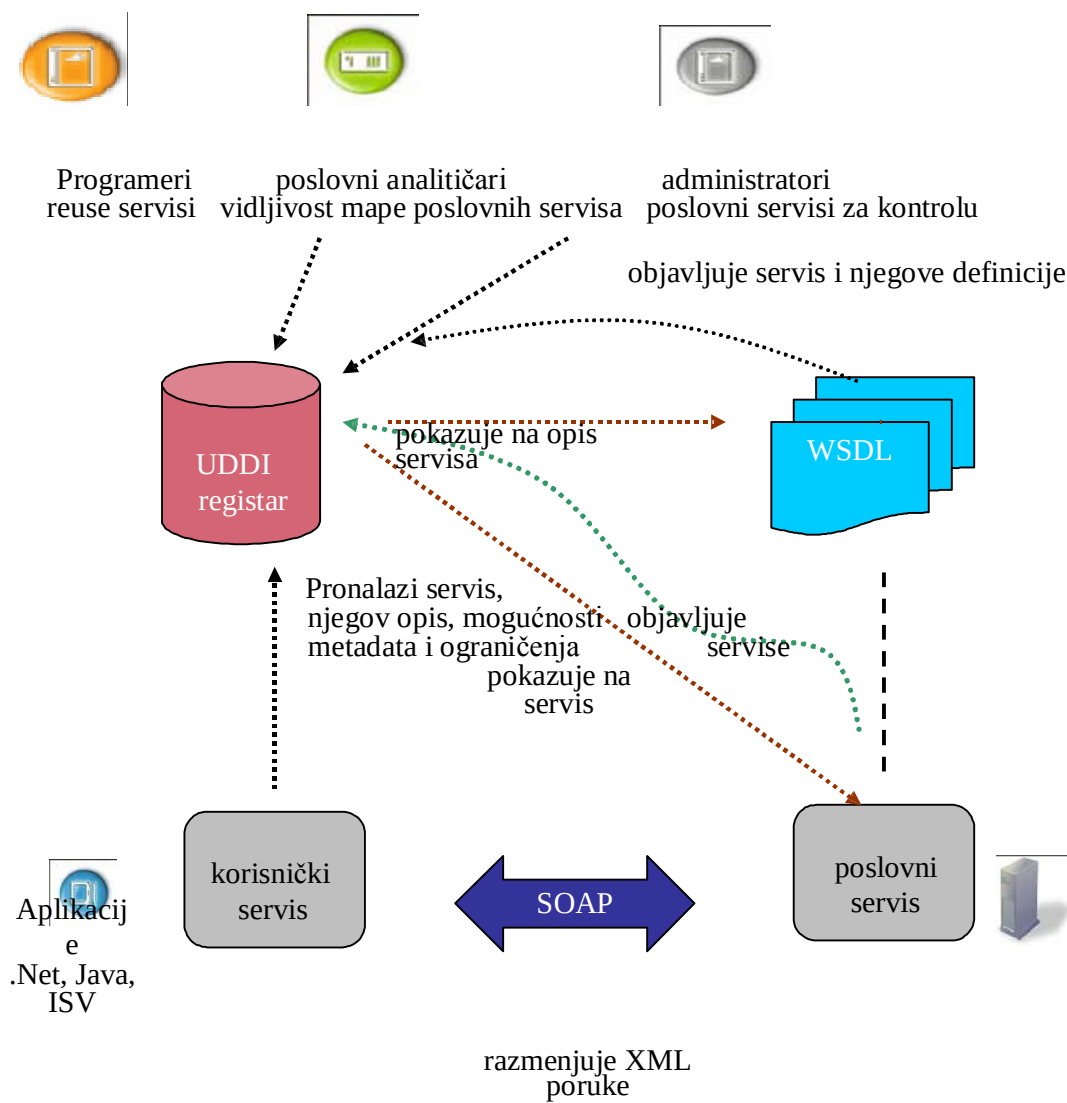
5.2 Osnovne karakteristike UDDI-a

UDDI je akronim za Universal Description, Discovery, and Integration To je nezavisno-platformski, XML bazirani registar za poslovne kompanije sa ciljem njihovog predstavljanja na Internetu. UDDI omogućuje kompanijama da objavljuju liste servisa i da se međusobno pronalaze. On definiše kako servisi softverskih aplikacija međusobno utiču jedni na druge preko Interneta.

UDDI poslovni registar čine tri komponente:

- ✓ Bele strane – adresa, kontakt, ime (naziv);
- ✓ Žute strane – industrijske kategorizacije zasnovane na standardnim klasifikacijama;
- ✓ Zelene strane – tehničke informacije o servisima izloženim od strane kompanija.

UDDI je jedan od osnovnih standarda Web servisa. Dizajniran je da bude upitnik za SOAP poruke i da omogući pristup WSDL dokumentima, opisujući obavezne protokole i formate poruka potrebne za interakciju sa Web servisima navedenim u direktorijumima određenih dokumenata. UDDI je napisan avgusta 2000. godine, u vreme kada su autori imali viziju sveta u kome bi korisnici Web servisa bili povezani sa provajderima kroz javni ili privatni dinamički posrednički sistem. U ovoj viziji, svako kome bi bila potrebna određena usluga, kao na primer provera verodostojnosti kreditne kartice, bi otišao na njihov posrednički sistem i izabrao jedan od podržavajućih, SOAP ili drugi servisni interfejs. Na taj način, javno (otvoreno) rukovanje UDDI posrednikom bi bilo značajno za sve. Za korisnike, javni posrednički sistemi bi vraćali samo liste servisa koje su dostupne svima, dok za servisne dizajnere informacije o uspešnom nalaženju zaposlenja, oslanjajući se na meta podatke merodavnih kategorija. UDDI se može smatrati za projekat baziran na zakonskim propisima, u kome su Microsoft, IBM i Ariba nastojali da redefinišu kontrolu nad osnovnom mrežnom tehnologijom, preopterećujući tu posrednu tehnologiju preteranom složenošću. Ovi i kasniji finansijeri, imali su za cilj da daju UDDI na javno korišćenje kao i istraživanje servisa besplatno. UDDI je integrisan u tzv. WS-I „standard“ kao centralni stub arhitekture Web servisa. Do 2005. godine, u upotrebi je bio u više od 70 % od 500 najuspešnijih kompanija, privatno ili javno implementiran, naročito među onim kompanijama koje su pokušavale da optimizuju softver. Mnogi od ovih preduzeća su se pretplatili na neku od verzija SOA (Service Oriented Architecture), server programe ili softverske baze podataka, licencirane od strane priznatih osnivača UDDI.org i OASIS.



Slika Korišćenje UDDI registra, [4]

UDDI specifikacije podržavaju javni pristup opštom poslovnom registru (Universal Business Registry) u kome je ugrađen sistem imenovanja, na čemu sada finansijeri UDDI –a zahtevaju, u pogledu moguće odgovornosti, da je ovo bio potpuno nedokazan poslovni model. Neki tvrde da je najčešće mesto na kome se UDDI sistem može naći unutar kompanije, gde se koristi da dinamički poveže klient sisteme sa implementacijom. UDDI koristi pogodnosti WorldWideWeb Consortium (W3C) i Internet Engineering Task Force (IETF) standarda, kao što su XML (Extended Markup Language), HTTP i Domain Name System (DNS) protokoli.

Osnovni cilj pri stvaranju ove verzije je bio podržavanje sigurne interakcije private i public UDDI implementacija. UDDI v3.0 opisuje Web servise, strukturu podataka i ponašanje svih instanci UDDI registra. UDDI protokol je ključni član uzajamnih standarda koji obuhvataju mnoštvo Web servisa. Ova verzija je dizajnirana tako da se koristi u kombinaciji sa drugim dopunskim Web servis specifikacijama. UDDI v3.0 je razvijena na konceptu „meta servisa“ za lociranje Web servisa omogućavajući jasne upite, nasuprot kompleksnim metadata.

Proširujući osnovu OASIS UDDI v2 standarda, UDDI v3 nudi specifikaciju za razvoj fleksibilnih, interoperabilnih XML Web servis registara, upotrebljivih kako u public, tako i u private razvoju. Jedna od karakteristika UDDI3.0 je i mogućnost definisanja sopstvenih ključeva. UDDI V3.0.2 je odobren od OASIS organizacije februara 2005. godine. Smatra se kamenom temeljcom Web servisa.

Definiše standardne metode za objavljivanje i pronalaženje mrežno baziranih softverskih komponenti u servisno orijentisanoj arhitekturi SOA (Service Oriented Architecture). Razvijen je u okviru otvorenog procesa. Podržava pripajanje registra.

Bitne karakteristike sa pogleda izdavaštva su da je jednostavan za upotrebu, URI baziran, podržava API, digitalne potpise (što omogućava da UDDI poboljša stepen ispravnosti i nezavisnosti podataka), poboljšane su karakteristike pretrage (sa složenih upita pređeno je na jednostavne, gde podupiti mogu biti ugnježdjeni u jednostavne singl upite), promena u osnovnom mehanizmu od pokazivača – smernica do konkretnih servisa, od Globally Unique Identifiers (Globalnih Jedinstvenih Identifikatora) do Web baziranom Universal Resource Identifier (Univerzalnom Izvornom Identifikatoru). UDDI v3.0.2 bolje ispunjava zahteve preduzeća od prošlih implementacija u pogledu rasta mehanizama poverenja, jačom kategorizacijom i bezbedonosnim karakteristikama.

Definiše standardne metode za objavljivanje i pronalaženje mrežno baziranih softverskih Komponenti u servisno orijentisanoj arhitekturi SOA (Service Oriented Architecture). Razvijen je u okviru otvorenog procesa. Podržava pripajanje registra. Bitne karakteristike sa pogleda izdavaštva su da je jednostavan za upotrebu, URI baziran, podržava API, digitalne potpise (što omogućava da UDDI poboljša stepen ispravnosti i nezavisnosti podataka), poboljšane su karakteristike pretrage (sa složenih upita pređeno je na jednostavne, gde podupiti mogu biti ugnježdjeni u jednostavne singl upite), promena u osnovnom mehanizmu od pokazivača – smernica do konkretnih servisa, od Globally Unique Identifiers (Globalnih Jedinstvenih Identifikatora) do Web baziranom Universal Resource Identifier (Univerzalnom Izvornom Identifikatoru). UDDI v3.0.2 bolje ispunjava zahteve preduzeća od prošlih implementacija u pogledu rasta mehanizama poverenja, jačom kategorizacijom i bezbedonosnim karakteristikama.

5.3 Protokoli za registre web servisa

1. Standard za vidljivost i ponovnu upotrebu SOA komponenata
2. Standard za prilagodljiva preduzeća–dinamičko otkrivanje i vezivanje za SOA

Tipična aplikacija za registre omogućava objavljivanje ili pronalaženje Web servisa (u okviru organizacije) na osnovu proizvoljnog kriterijuma. Takođe omogućava da se definišu bezbednost i transport protokola podržanih određenim Web servisom, kao i funkciju ogradaivanja aplikacija od neuspeha ili promena u servisima koji se pozivaju.

5.4 Karakteristike UDDI-a sa aspekta arhitekture

UDDI servisi predstavljaju interoperabilno rešenje koje omogućava da se Web servisi pronađu i upotrebe sa različitih razvojnih platformi u okviru date organizacije. Interoperabilna priroda UDDI-a predstavlja mogućnost da svi programeri jedne organizacije, bez obzira na jezik ili razvojno okruženje, mogu da iskoriste UDDI servisni registar. Iznad svega, UDDI servisi su omogućili design-time reuse Web servisa, u njihovoj širokoj, heterogenoj sredini. Faze procesa date su kroz jedan primer:

1. Tim u okviru Qwest-a piše određeni Web servis.
2. Kada je Web servis spreman za izvođenje, korišćenje, član tima objavljuje Web servis u UDDI registar.
3. Kada su drugom timu u okviru Qwest organizacije potrebne informacije o novim zadacima, član tima pretražuje UDDI servise u potrazi za njima bitnim Web servisom.
4. Nakon otkrivanja određenog Web servisa pomoću UDDI-a, programer može da doda kod u projekat i tako pozove Web servis. Informacije o tome koga treba kontaktirati za više informacija o određenom Web servisu, takođe su obezbeđene. Lynn Fischer, šef katedre za informatiku i potpredsednik Qwest-a, je izjavio: "UDDI servisi su ključna komponenta Qwest IT strategije za implementaciju arhitekture Web servisa. Omogućavajući registraciju i pretragu Web servisa, pružaju mogućnost ponovne upotrebe (reuse) i eliminišu ponavljanja, a samim tim smanjuju troškove uz obezbeđivanje brže isporuke".

5.5 UDDI Produkt zahtevi

Da bi se UDDI mogao koristiti, bilo je potrebno da ispuni određene uslove, tj. zahteve:

- Moderan grafički korisnički interfejs (GUI) - Zbog obima interakcije sa UDDI registrom, za Qwest je bilo bitno da GUI za UDDI bude moderan, ali istovremeno i jasan i jednostavan za razumevanje.

- Interoperabilnost – Zbog raazličitosti platformi u okviru Qwest organizacije , UDDI je morao da zadovolji određeni stepen interoperabilnosti. Softver je morao da povezuje različite UDDI klijente, dostupne na različitim platformama.
- Osobina podržavanja standarda – Za Qwest je bilo važno da UDDI produkt bude u saglasnosti sa verzijom 1 i verzijom 2 UDDI specifikacija.
- Step en učinka – Kako UDDI obezbeđuje ključne informacije za Qwest Web servis infrastrukturu, tako je UDDI produkt morao da se izvršava sa visokim stepenom pouzdanosti i korisnosti.
- Bezbednost – Zbog činjenice da UDDI registar sadrži bitne, ključne informacije, Qwest je zahtevao jaku sigurnosnu infrastrukturu za postizanje autentičnosti UDDI-a.

Stvorili su sistem u kome korisnici imaju različit stepen prava u korišćenju registra (crveni, beli,...).

5.6 Modelovanje podataka

Jedan od prvih koraka za bilo koju organizaciju, u razvijanju UDDI servisa, je praksa u radu sa UDDI data modelom i razumevanje načina na koji će organizacija moći da mapira svoje strukture i procese, u skladu sa UDDI specifikacijom. Razvojni tim je želeo da razume i što više iskoristi način na koji sistem prikazuje podatke. Qwest je pažljivo razmatrao način na koji bi Web servisi mogli da se modeluju u UDDI. Tako je uspostavljena strategija procene podataka, da bi se obezbedili da preporuke modelovanja podataka koje ustanove, budu dosledno ispraćene.

5.7 Modelovanje četiri osnovna elementa UDDI-a

Tri od četiri entiteta, provajderi, servisi i veze, imaju jasnu, hijerarhijsku vezu. Servisi su logični naslednici („deca“) provajdera, a veze naslednici servisa. Preostali entitet, poznat kao tModel, egzistira izvan ova tri entiteta i ponaša se kao deljena struktura preko drugih entiteta. Jedna od prvih odluka koje se moraju doneti kada je reč o modelovanju podataka u UDDI-u , je definisanje šta predstavlja provajder u organizaciji. U zavisnosti od situacije, provajder može da se modeluje na više različitih načina. Na najapstraktnijem nivou, provajderi pružaju servise. Zatim se postavlja pitanje identifikacije provajdera u Qwest-u. Dolazi se do zaključka da svaki projekat Web servisa, koga čine programeri, menadžeri i tester i, u stvari predstavlja provajdera. Nakon te odluke ostatak modela dolazi na svoje mesto. Web servisi koji su razvijeni kao rezultat projekta postaju servisi koje provajder nudi. Tačke pristupa u kojima se pozivaju ovi Web servisi postaju veze servisa. Na kraju se Web Service Description Language (WSDL) dokumenti u vezi sa servisima predstavljaju kao tModel elementi.

5.8 Strategija kategorizacije

Bitan aspekt UDDI data modela je sposobnost kategorizacije ulaza korišćenjem jedne ili više klasifikacionih šema. Provajderi, servisi i elementi tModela mogu biti klasifikovani, tako da ih je prilikom pretraživanja lakše naći. Jeff Kenyon, član Architecture and Transversal Services

grupe je izjavio: "Ispravna kategorizacija ulaza UDDI servisa bila je ključna stvar za postizanje sposobnosti da prikupimo i ujedinito korisne strane Web servis pristupa razvoju aplikacije. Ukoliko neki ulaz nije kategorizovan, šanse za njegovo pronalaženje i korišćenje su jako male. Ključne reči pretrage su nepouzdanost određene korišćenjem različitih sinonima, različitog modela između razvojnog tima i korisnika i td., tako da je neosnovano očekivati da korisnik pretražuje čitav katalog servisa u nadi da će pronaći nešto korisno." [6]

5.9 UDDI kontrole i procesi

Qwest je nakon više razmatranja odlučio da razvojna sredina bude sa višestrukim registrima. Svaki registar služi za različite svrhe, a podaci se prevode i usklađuju između registara, ukoliko je to potrebno. Tako su stvorena četiri UDDI servisna registra, svaki sa jasnom svrhom i korisničkom grupom.

1. Registar objavljivanja – Ovaj registar služi kao oficijalni zajednički Qwest registar. Za web servise koji se pronadu u ovom registru garantuje da su u potpunosti testirani, spremni za upotrebu od strane klijenata u organizaciji. Mogućnost objavljivanja je ograničena na manji skup individua.

2. Test registar – Ovaj registar služi kao ogledalo registru objavljivanja. Ulazi se dodaju test registru pre nego se krene na izradu, tako da se na datim ulazima kontrola kvaliteta može izvršiti pre njihovog prebacivanja u registar objavljivanja. Prava istraživanja su široko garantovana, tako da različiti test timovi mogu da pristupe podacima, ali prava izdavanja (objavljivanja) su ograničena na manji broj individua.

3. Registar razvoja – Ovaj registar je dostupan za petraživanje i objavljivanje svim Qwest razvojnim timovima. Razvojni timovi moguda popune ovaj registar po volji, Webservisima bilo koje vrste, kvaliteta ili složenosti. Kada razvojni tim smatra da je određeni Web servis spreman za objavljivanje, ulaz se prebacuje u test registar radi kontrole kvaliteta testiranjem. 4. Registar „Peščana kutija“ (Sandbox) - Ovaj registar u potpunosti služi za eksperimentisanje, kao i testiranje softverskih novina. Ključna stvar kod topologije višestrukih registara je pravilna razmena podataka između registara. Qwest je ustanovio proces koji pažljivo kontroliše način kako se ulazi registra prebacuju sa jednog registra na drugi. Kada je Web servis spreman za objavljivanje, član razvojnog tima razvija odgovarajuće ulaze u registru razvoja. Na primer, ulazi mogu biti tModel elementi koji predstavljaju WSDL dokument Web servisa i provajdera, servisa i veza koje se pristupna tačka Web servisu. Zatim se verifikuje da ulaz odgovara UDDI standardu Qwest organizacije. Tvorac projekta takođe vrednuje (procenjuje pravosnažnost) ulaza registra. Kada su POC i tvorac projekta zadovoljni ulazima registra i kada ne dođe do ispoljavanja ikakvih grešaka, oni šalju zahtev kroz e mail odgovarajućem aliasu. Poruka uključuje nazive provajdera, servisa i tModel elemenata koji se prenose i zahtev za prebacivanjem ulaza registra. Nakon što je Web servis razvijen i testiran, pojedinac imenovan kao AIP (Application In Production) šalje e mail poruku odgovarajućem aliasu, zahtevajući da ulaz registra bude prebačen u registar objavljivanja. Poruka takođe sadrži nazive provajdera, servisa i tModel elemenata koji se prenose. Ovim procesom se garantuje da ljudi iz razvojnog tima imaju kompletno znanje o Web servisima, iz svoje oblasti. Tako se omogućuje garancija da se samo odobreni Web servisi iz njihove oblasti registruju.

5.10 Validacija podataka

Da bi se obezbedila ispravnost podataka, izvršava se sintaksna i semantička provera podataka, pre prebacivanja ulaza u test registar.

5.11 Sintaksna provera

Sintaksna provera se odnosi na procenu da li ulazi imaju određena polja (oblasti) kompletno završena. Neka od pravila koja je Qwest ustanovio, a koja se moraju ispoštovati su:

- Svaki ulaz UDDI servisa mora da ima opis.

- Svaki ulaz UDDI servisa mora da bude kategorizovan TOM klasifikacijom.
- Automatski generisani WSDL fajlovi se ne bi trebali koristiti kao tModel elementi.

Zbog toga što se ova pravila oslanjaju na normalizovane sintaksne odredbe Architecture and Transversal Services grupe, ulazi mogu da se proveravaju na programskom nivou koristeći UDDI API. Qwest je razvio interni alat za proveru ulaza, koristeći Java toolkit.

5.12 Semantička provera

Alati za sintaksnu proveru koje je Qwest razvio pomažu u automatizaciji procesa validacije, ali njima se ne može u potpunosti izvršiti provera ulaza. Neke smernice zahtevaju ljudsko tumačenje. Semantička provera se izvršava na svakom ulazu određenom za prebacivanje. Ova semantička provera odnosi se kako na same ulaze UDDI servisa, tako i u slučaju tModel elemenata na WSDL fajl koji predstavlja tModel element. U ovoj proveru, obezbeđuje se sledeće:

- Nazivi su kratki, jednostavni, smisleni i opisni.
- Opisi su koncizni, smisleni i potpuni.
- Nazivi Web servisa su imenice, au slučaju kompleksnosti naziva, prvo slovo svake interne reči je veliko (na primer, DSLOrderingService) ili sa razmakom između svake reči (DSL Ordering Service).

Sledeće provere se izvršavaju na WSDL fajlu:

- Nazivi metoda su glagoli, u složenim slučajevima sa prvim velikim slovom svake reči. Na primer, DSLOrderingService može da ima placeOrder metodu.
- Nazivi parametara metoda su kratki, ali smisleni. Nazivi parametara kao što su x,y,z ne bi trebalo da se koriste.
- Koriste se samo cele reči; akronime i skraćenice treba izbegavati, sem u slučaju kada je skraćenica više u upotrebi od pravog (dužeg) naziva. Samo ulazi koji su prošli validacionu proveru mogu da prelaze iz registra u registar. u suprotnom, potrebno je ispraviti greške.

5.13 Modelovanje ne web servisa u UDDI servise

Qwest istražuje kako drugi tipovi servisa mogu da se modeluju u UDDI servise. Ovi servisi uključuju druge tipove RPC (Remote Procedure Call) servisa u okviru svog intaneta, kao što su CORBA (Common Object Request Broker Architecture) elementi, interne softverske aplikacije i servisi bazirani na ljudskoj delatnosti. UDDI data model može da se prilagodi na niz različitih tipova servisa. Qwest je stvorio jedno centralno skladište gde se čuvaju svi servisi u okviru organizacije, tako da mogu da se pronađu kroz niz različitih značenja. „Radimo na tome da proširimo delovanje našeg UDDI servisnog direktorijuma sa Web servisa na celu našu softversku imovinu, kao što su database servisi, BIM (Business Information Model) i drugi sistemi. Cilj je da obezbeđujući centralni UDDI servisni registar olakšamo upotrebu i bolju optimizaciju našeg softvera“, rekao je Mahesh Dalvi, direktor u Architecture and Technology Services u Qwest-u. [6]

5.14 UDDI API

Postoji mnogo definisanih API u UDDI-u koji se koriste kako u upitima, tako i u objavljivanju. UDDI API upiti podržavaju nekoliko tipova obrazaca upita: 1. „Browse“ obrazac omogućava široku pretragu korišćenjem „find_xxx“ API –a. 2. „Find_xxx“ API omogućava korisniku da pretraži registar na osnovu više kriterijuma pretrage. 3. „Drill-down“ obrazac se bazira na „get_xxx“ API koji sadrži kompletne detalje o odgovarajućem ulazu. 4. „Invocation“ obrazac preuzima korisnike iz memorije registra vezujući informacije za servise koje oni koriste, a u registar vraća samo kad je neophodno ažuriranje. „Save_xxx“ and „delete_xxx“ komande API omogućavaju kreiranje i uništavanje bilo kog entiteta. „Get_authToken“ and „discard_authToken“ API komande se koriste za identifikaciju verodostojnosti uverenja za objavljivanje u fazi priprema za objavljivanje entiteta.

5.15 Primer primene UDDI-a u realnom sistemu

UDDI registar pomaže NASI u stvaranju svetske naučne „pijace“ Današnji biznis zahteva softver agilnije arhitekture koji omogućava brže pronalaženje i korišćenje IT. Pre nego što su akronimi kao SOA, SOAP i UDDI postali deo naučne terminologije, svetska naučna organizacija pod pokroviteljstvom NASA, počela je sa razvojem SOA za Earth Observing System Clearinghouse (ECHO) koji omogućava naučnicima da pristupaju, pretražuju i dele svetske naučne podatke. NASA je sada stvorila prošireni servisni registar za ECHO, baziran na UDDI registru koji omogućava trećim licima da objavljuju i pristupaju podacima. ECHO, započet 1998. godine, koristi XML i Web tehnologiju. Systinet UDDI server se koristi da izloži servisni registar kroz standardni UDDI protokol. ECHO se dorađuje da bude kompatibilan sa Web Services Interoperability (WS-I). U verziji 8, ECHO postaje dostupan kao Web servis.

Frank Kenney, analitičar iz Gartner istraživačkog centra je izjavio: „Ogromna je moć SOA uticaja kada se procesi, polise, SLA i druge informacije u vezi sa servisima skladište u registar. Verujemo da će registri biti od ključnog značaja za otkrivanje i skladištenje servisa, sa prednošću omogućavanja funkcije nadgledanja.“ [7]

UDDI je na mala vrata ušao u svet Web servisa i vremenom postao njegov nezamisliv deo. Sada predstavlja budućnost Web servisa u koju se sve više ulaže i to sa razlogom. 16.

Ovaj deo rada sadrži priču o UDDI-a, njegov razvoj kroz vreme, definicije, standarde, protokole zaregistrovanja Webservisa, karakteristike UDDI-a sa aspekta arhitekture, osvrt na proces modelovanja podataka pri razvoju UDDI servisa kao i proces validacije podataka, pregled UDDI kontrola i procesa, modelovanje servisa koji nisu Web bazirani u UDDI servise kao i primer praktične primene UDDI-a u realnom svetu.

6. Pretraživači

Svet je postao veliko tržište, a s obzirom da živimo u eri informatike i moderne komunikacione tehnologije, prava i tačna informacija znači sve: zauzeti što bolje mesto i što duže opstati na tržištu. Pojavom globalne mreže veliki broj informacija je postao dostupan široj populaciji, čime se i konkurencija na tržištu zaoštrava jer je mnogo lakše pratiti trendove i prilagođavati svoj proizvodni asortiman korisnicima. Naravno, tu su i korisnici koji Internet koriste za zabavu i kontakt sa rođacima i prijateljima iz udaljenih mesta, što doprinosi da Internet postane mreža globalnih razmera, a ujedno i naša svakodnevna potreba. Zbog toga javila se potreba za servisom koji bi mogao da umesto korisnika pretraži postojeće stranice i u mnogome olakša potragu i dolazak do željenih informacija. Iz ove potrebe dolazi do pojave i razvoja search engine-a bez koga bi danas bilo nemoguće koristiti moć Interneta.



7. Teoretske osnove pretrage Interneta

7.1 Pretraživanje WWW

Neverovatan rast količine informacija ponuđenih na web-u putem Interneta otvara pitanje snalaženja i vrednovanja ponuđenog materijala. Postoje dve osnovne tehnike pretraživanja:

- Pomoći ključne reči
- Kataloški

Pretraživanje pomoću ključne reči zasniva se na zadatom skupu pojmova. Pretraživački program pretražuje bazu podataka web adresa tražeći one koje sadrže ključne reči. Ukoliko adresa sadrži tražene pojmove program nas obavesti o adresi i povećava brojač adresa za jedan. Brojač nas obaveštava o broju adresa sa traženim pojmovima, ukoliko je taj broj prevelik (100000) upit nije dobar jer je prevelik, u tom slučaju dodajemo pojam koji će suziti izbor. Najpoznatiji takvi pretraživači su Google, Yahoo i dr.

Drugi način traženja je prolazak kroz web katalog putem područja. Prikazano stablo pojmova koje nas vodi do traženih web adresa. Ovaj način se zasniva na poznavanju područja. Moramo unapred poznavati koji niz pojmova nas vodi do tražene teme. Moguće je prolaziti niz stablo i metodama eliminacije ili pokušaja, a možemo jednostavno koristiti i intuiciju.

7.2 Search engine

Search engine je servis koji nam pomaže da na Internetu lakše pronađemo sadržaje koje tražimo. Pomoću search engine-a možemo pronaći sve što nas zanima na Internetu, na raznim web stranicama, među brojnim e-mail adresama u svetu i slično. Delovi search engine-a su polje za upisivanje izraza koji želimo naći, „pull down meni“ za određivanje opsega pretraživanja i dugme za početak pretraživanja.

7.3 Napredno pretraživanje Interneta

Ukoliko želimo da dobijemo što tačnije rezultate pretraživanja, treba da koristimo napredno pretraživanje. Kada u pretraživač upišemo samo jednu reč, dobićemo sve stranice koje sadrže tu reč, a kada upišemo izraz od više reči, dobićemo stranice koje sadrže bilo koju od upisanih reči. Da bi precizirali šta tražimo, možemo koristiti mogućnost naprednog pretraživanja. Od specijalnih znakova možemo upotrebiti: ", +, -, *, i, ?.

- Upotrebom navodnika kod upisivanja više reči search engine će pronaći samo stranice koje sadrže reči koje smo upisali i to tačno istim redosledom
- Upotrebom znaka „+“ search engine će pronaći sve stranice koje sadrže te reči, ali bez obzira na redosled
- Upotrebom znaka „-“, search engine će pronaći sve stranice koje sadrže neke od reči
- Upotreba znaka „*“ nam omogućava da upišemo deo neke reči, a ostalo zamenimo znakom „*“
- Upotreba znaka „?“ zamenjujemo bilo koji znak ili slovo u izrazu koji tražimo

Postoji još i pretraživanje pomoću Booleove algebre, logičkog sistema koji je u 19. veku predstavio matematičar George Boole. Tri osnovna logička operatora predstavljaju tri engleske reči AND, OR, NOT koje služe za stvaranje sintakse naprednog pretraživanja.

AND - koristimo kako bi smo pretraživaču naredili traženje dokumenata u kom se nalazi više ključnih reči.

OR - kao rezultat donosi najjednostavnije odgovore pa je zbog toga postavljen kao osnovni logički simbol većine pretraživačkih sistema.

NOT - isključuje određene reči iz traženog niza, tj. funkcija mu je kao kod znaka „-“

8. Pretraživanje Interneta

Od dana kada je nastao, Internet se znatno promenio. Internet možemo opisati na razne načine, ali posle svega ostaje jedno: da je to jedan odgroman, nepresušan izvor podataka i ako ga naučimo koristiti, poboljšaćemo kvalitet života, a kako je i on odlična alatka za učenje i izvor svežih informacija, pomoću njega možemo dobiti veoma bitne informacije.

8.1 Osnove pretraživanja

Da bi se došlo do određenih informacija na Internetu, koriste se pretraživački alati (search tools). Oni rade tako što korisnik unese pojam koji traži, nakon čega dobije listu adresa na kojima je taj pojam pronađen. Na Internetu postoji veliki broj pretraživačkih alata, koji mogu biti opšti ili specijalizovani za pretragu dokumenata iz samo pojedinih oblasti, a generalno se svrstavaju u dve kategorije:

- pretraživači
- katalozi

Postoji i treća vrsta web sajtova, koji se nazivaju metapretraživači. Oni su obično sporiji od kataloga i pretraživača, ali kombinujući ih sa više pretraživačkih mašina, pokrivaju veći procenat Interneta, pa zato mogu dati bolje rezultate kada se traže stvari koje nisu toliko poznate. Osim pretraživača postoje i programi koji vrše istu funkciju i rade na sličan način, a najpoznatiji program takve vrste je Copernik.

Prilikom pretraživanja neke reči koje treba uključiti u pretragu su sledeće:

REVIEW - ako tražite pregled nekog proizvoda
DOWNLOAD - ako tražite neku datoteku koja je stavljena za skidanje sa Interneta
GALLERY - ako tražite neke slike
PRICE - ako tražite cenu nekog proizvoda
COMPARE - upoređivanje nekog proizvoda
SPECIFICATION - dobićete link na tehničke specifikacije nekog proizvoda

8.2 Strategija pretraživanja

Ako ste traži adresu na Internetu moraju se navesti reči koje nas interesuju. Kada pretraživač nađe reči koje su mu zadate, kao rezultat izlistaće niz linkova sa adresama na kojima se nalaze tražene reči. Kada se pretrazuje Internet, korisnik mora sam da uradi filtriranje informacija. Pretraživanje Interneta možemo svesti na nekoliko faza:

- Prvo što se radi je da se ode na neku od stranica koje sadrže pretraživač, zatim se u polje za unos upišu ključne reči koje pretraživač želimo da nađe.
- Nakon što je pretraživanje izvršeno, drugo što se radi je probiranje kroz linkove koje dobijemo kao rezultat pretrage. U ovoj fazi se vrši filtriranje informacija.
- Treće što se radi je pregled stranica koje smo odredili i konstatovanje da li smo našli ono što smo tražili

8.3 Napredna pretraga

Na stranici za naprednu pretragu se može preciznije definisati pojam koji tražimo. Pošto je u bazu uključeno preko trideset i pet ne HTML dokumenata, pretraga se može ograničiti i na neki od njih. Takođe pretraga se može ograničiti i po datumu, tako što odredimo u rezultatima da prikaže samo web stranice u poslednja tri meseca ili godinu dana i sl., odnosno odredimo da se traženi termini nalaze bilo gde na stranici, ili u linkovima koji vode do nje. Još jedna od zanimljivih pretraga, je pretraga unutar stranice odnosno domena. Pomoću ove pretrage može se uraditi i suprotno, zabraniti prikaz rezultata sa određenog sajta odnosno domena. Postoje dve zanimljive mogućnosti: da pretraživač prikaže sve slične stranice odnosno sve stranice koje imaju link ka nekoj stranici i može se koristiti za pronalaženje većeg broja sajtova koji pišu o nekoj temi. Takođe postoji i način da sve ove pretrage izvršimo direktno iz polja za pretragu korišćenjem određenih operatora. Operatori za naprednu pretragu su: LINK, INFO, SITE, ALLINTITLE, INTITLE; ALLINURL i FILETZPE. Između operatora i reči koja ga sledi ne ostavlja se prazan prostor.

The screenshot shows the 'Advanced Search' interface. At the top, there's a header with 'Advanced Search' on the left and a link to 'Advanced Search Tips' on the right. Below the header is a light blue box with the instruction: 'Use the form below and your advanced search will appear here'. The main form is divided into several sections. The first section, 'Find web pages that have...', contains three input fields: 'all these words:', 'this exact wording or phrase:', and 'one or more of these words:'. The second section, 'But don't show pages that have...', has an input field for 'any of these unwanted words:'. The third section, 'Need more tools?', includes dropdown menus for 'Reading level:' (set to 'no reading level displayed'), 'Results per page:' (set to '10 results'), 'Language:' (set to 'any language'), and 'File type:' (set to 'any format'). A note states 'This option does not apply in Google Instant'. There is also a text input for 'Search within a site or domain:' with an example '(e.g. youtube.com, edu)'. At the bottom left, there's a link for 'Date, usage, rights, numeric range, and more'. An 'Advanced Search' button is located at the bottom right of the form.

©2011 Google

8.4 Pretraga slika

Pretraživači takođe nude i mogućnost pretrage slika. Kada unesemo željeni pojam, dobićemo prikazane umanjene verzije indeksiranih slika. Klikom na neku od njih, ulazimo na web stranicu na kojoj je ta slika upotrebljena. Takođe postoji i napredna pretraga slika, koja omogućava da odredimo veličinu slike koja nas interesuje kao i boju ali isto se može i ograničiti pretraga na određeni domen.

9. Razlika između web direktorijuma i search engine-a

Danas je između njih manja razlika nego ranije, jer je veliki broj search engine-a izradio velike kataloge u nameri da olakšaju našu pretragu.. Yahoo pokušava da organizuje web (indeksirane stranice) deleći ga na teme i pod teme: umetnost, nauka, posao, zabava... Ukoliko Internetom tragamo za nekom informacijom iz opštih kategorija potrebno je prvo otići na Yahoo ili neki drugi web direktorijum. Ako web direktorijum posmatramo kao indeks koji nam omogućava da pronademo posebnu reč ili frazu, kao rezultat pretrage dobijamo previše rezultata, ili možemo da otkrijemo da reč kojom pretražujemo ima neko sasvim drugo nama nepoznato značenje.

Yahoo - je zapravo web direktorijum sa indeksiranim temama. To je pokušaj katlogizacije celokupnog web-a, pretražujemo ga po predmetu ili temi.

Google - je jedan od najpoznatijih search engine-a koji je trenutno najomiljeniji sajt za pretragu. Google je veoma temeljan i brz, ali pravilo mu je da daje više važnosti stranicama koje su povezane sa drugim stranicama, a manje važnosti daje stranicama sa malim brojem linkova.

10. Serach engine

Opšta podela pretraživača je izvršena na osnovu načina indeksiranja strana. Ako je neka strana indeksirana od strane pretraživača, to znači da je njen deo ili ceo sadržaj ubačen u bazu tog pretraživača i da će iz nje biti pozvan, prilikom pretrage ako sadržaj odgovara indeksiranom. Iz toga je nastala podela na:

- Web direktorijume
- Krolerske ili spider pretraživače
- Meta pretraživače

10.1 Web direktorijumi

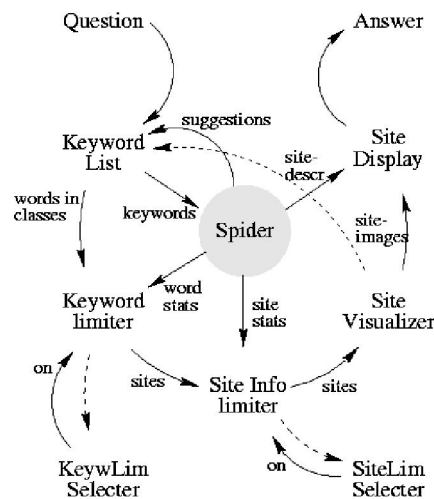
Web direktorijumi predstavljaju, uslovno rečeno najprimitivniji servis za podršku korisnicima prilikom pretrage interneta. To je prva vrsta search engine-a, i kao takva je prevaziđena u mnogim stvarima. Indeksiranje na web direktorijumima vrši sam čovek ili u većem slučaju više ljudi, odnosno fizički se u kategorije (direktorijume) ubacuje adresa određenih sajtova.

Mana ovog načina indeksiranja je očigledna, neophodni su ogromni ljudski resursi da bi svojom bazom mogli da pariraju ostalim vrstama pretraživača. To nije isplativo pa su baze ovih pretraživača uglavnom male i najčešće se koriste kao specijalizovani pretraživači. Nikada se ne indeksira ceo sadržaj site-a već samo ključne reči, naslovi i reči u domenu.

Prednost ovih pretraživača je u tome što čovekov faktor utiče na sadržaj direktorijuma tako da je gotovo nemoguće da za određenu pretragu doobijete nešto što nema veze sa tom reči, što je ipak moguće kod krolerskih pretraživača. Ovo je i jedini razlog zašto se ovi pretraživači i dalje koriste.

10.2 Krolerski ili spider pretraživači

Način indeksiranja ovih pretraživača je mnogo napredniji. Pretraživači su razvili posebne programe (krolere ili spidere) koji sami traže i neverovatnom brzinom indeksiraju ceo sadržaj stranice, odnosno site-a, čime se baza znatno povećava. Kroleri rade po utvrđenom algoritmu koji se može menjati.



Prednosti ove vrste indeksiranja su mnogo veća baza od baze web direktorijuma. Revolucionarno što su doneli kroleri je da je indeksiranje baš svake reči na strani i svake strane na sajtu, tako da danas uz pomoć ovakvih pretraživača gotovo da ne postoji reč za koju nećete

dobiti link kao rezultat pretrage. Pošto savršenstvo ne postoji, i kod ovog pretraživača imamo mane koje su u zavisnosti od same organizacije i mogućnosti ljudi koji rade na njemu. Jedna od mana je što uvek postoji mogućnost da kao rezultat pretrage dobijete nešto što nema veze sa pretragom.

10.3 Meta pretraživači

Pitanje je da li ovakav pretraživač uopšte možemo nazvati pretraživačem, jer u najvećem broju slučajeva ne poseduje svoju bazu već koristi tuđe, najčešće baze krolerskih pretraživača. Odnosno omogućava nam da pretražimo pomoću naprednog pretraživača a da zaobidemo njegov interfejs, što i nije neka prednost, ali jeste to što meta pretraživači najčešće koriste baze od više pretraživača, tako da je veća verovatnoća da za određenim traženjem dobijemo rezultat. Međutim, tu se javlja i problem dupliranja podataka, zbog čega je potrebno formatirati određene filtere koji će taj problm rešavati.

10.4 Problemi pretrživača

Od pojave Interneta do danas web masteri se trude da budu što bolje rangirani na listi pretraživača, jer neka ispitivanja govore da korisnik ide najčešće na prvi, eventualno drugi ili treći link. Ljudi koji prave pretraživač, prave ga da bude od koristi i da bude efikasan kako bi stekao poverenje i posećenost a samim tim i profit, dok je web masterima u intresu da privuku korisnike Interneta na svoj sajt i ako on nema veze sa potrebom korisnika. Zbog toga su i kroleri imali svoj razvitak od najjednostavnijih do ovih trenutno aktivnih koji uspevaju da prevaziđu većinu problema. Najveći problem današnjih pretraživača jeste ogroman broj podataka i informacija i kako ih sortirati, rangirati. Cilj je napisati pretraživač koji će svoje podatke u bazi sortirati na način na koji će biti od najveće pomoći korisnicima čak i u situaciji koja je vrlo česta, a to je da ni sam korisnik ne zna šta traži. Potrebno je napraviti algoritam koji će rangirati rezultate. U početku se rangiranje zasnivalo na tome koliko se puta tražena reč nalazi na strani, što više puta, to je bolje mesto na pretraživaču. Tada su web masteri došli na ideju da pišu ključne reči istom bojom kao što je i pozadina tako da se reči ne vide a kroler ih vidi i broji. Uskoro je i taj problem bio rešen, bar kod ozbiljnih pretraživača i takvi sajtovi su stavljani na crne liste. Sledeća stvar koja se javila, koja je za korisnike problem, a za pretraživače nije, jesu sponzori. Pretraživači koji su se okrenuli malo više ka profitu, su u rezultirane pretrage ubacivali i sponzore i to vrlo dobro kotirane, i ako nisu zadovoljavali pretragu korisnika, Takvi pretraživači su gubili poverenje i posećenost korisnika a samim tim i sponzore. Sledeći problem koji se javlja je dupliranje rezultata odnosno jedan korisnik - više domena. Korisnik ne zeli da na listi rezultata vidi na primer www.link.com i www.link.net ako su to iste prezentacije, problem je u tome koji prikazati a koji ne. Svaki pretraživač rešava te probleme na svoj način ali je google u tome najbolji.

11. Google

Za razliku od ostalih „.com“ kompanija, Google karakterise veoma uspešan, konstantan i progresivan stil razvoja. Kada je IT tržište, pogotovo područje Interneta bilo zasićeno došlo je do toga da su brojne kompanije bile ugašene, a paralelno sa tim Google se razvijao bez većih poteškoća. Glavna karakteristika Google-a i uzrok njegovog uspeha su neprestane inovacije, Većina usluga i proizvoda na kojima se temelji Google, plod su inovacija i ne konvencionalnog pristupa rešavanju problema. Google je postao kompanija za koju se može reći da dominira Internetom. Kada se spominju Internet pretraživači, programi, oglašavanja i distribucija medijskog sadržaja Internetom, Google ne samo da posluje u tom sektoru nego su najčešće i neophodan deo tog sektora.

11.1 Arhitektura

Google Inc. širom sveta poseduje više kompjuterskih centara, koji pojedinačno sadrže potpunu funkcionalnost pretraživača. Korisnički zahtev se kroz se kroz „Domain Name System“ (sistem koji imena google.com pretvara u IP adrese) prenosi u idealnom slučaju u sledeći mrežno topološki najbliži centar, navedeni centar je samo u retkim slučajevima identičan najbližem geografskom centru. U slučaju ako određeni kompjuterski centar nije sposoban za rad, preostali centri preuzimaju njegovu ulogu.

Svaki kompjuterski centar sastoji se od računarskog clustera. Za tu namenu korišćeni kompjuteri su personalni kompjuteri kompaktilni sa IBM-ovim proizvodima, prvenstveno kompjuterima. Navedeni sistem koristi samorazvijeni Google File System - razdeljena arhitektura koja sve podatke skladišti na različitim uređajima. U slučaju ako dođe do kvara određenog kompjutera ili određenog hard diska unutar kompjutera, odgovarajući podaci se sa drugog mesta unutar clustera prekopiraju na rezervni kompjuter, tako da se oštećena hardverska komponenta može zameniti u toku rada tj. prema potrebi mogu se dodavati ili oduzimati kompjuteri ili pojedinačni hard diskovi.

Odlaganjem određene strukture podataka (indeksi, kopije dokumenata, zapisi svake vrste) na različitim kompjuterima i zahvaljujući velikom broju raspoloživih kompjutera pojedini zahtevi mogu se relativno brzo obraditi, budući da se informacije čitaju paralelno. Vreme odgovora od maksimalno pola sekunde smatra se prihvatljivim, pa se koristi za orijentaciju kod utvrđivanja brzine kojom se isporučuju odgovori za zahtev. Brzo dobijanje informacija bio je veliki doprinos popularnosti Google-a već početkom razvoja sistema.

11.2 Software

Kao operativni sistem koristi se osiromašena verzija Linux distribucije Red Hat. Programski jezik C++ i Python imaju esencijalnu ulogu kod implementacije softvera pretraživača u celom sistemu.

11.3 Redosled stranica

Broj dokumenata koji odgovaraju zahtevu može biti nepregledne veličine. U tom slučaju potrebno je rezultate sortirati prema relevantnosti, kako bi se na zahtev odgovorilo sa najvažnijim dokumentima pozicioniranim na prvim mestima. Procena upotrebljivosti izvora u odnosu na zahtev jedan je od odlučujućih kriterijuma prema kojima se meri kvalitet

pretraživača. Na temelju količine za to određenih informacija dodeljeni Ranking se sa vremena na vreme automatski proverava pomoću algoritma. Proces rangiranja stranica se mora obaviti u razumnom vremenskom okviru, kako bi ranking bio aktuelan.

Google za organizaciju traženih rezultata među ostalima koristi i takozvani PageRank vrednost. PageRank vrednost se određuje prema broju linkova na određeni dokument. Koliko više postoji linkova, i koliko je veći Page Rank stranice na kojoj su smešteni linkovi, toliko je veća vrednost odgovarajućeg dokumenta. Pojedinačni PageRank stranice može se odrediti pomoći browser-plugin-a pod nazivom Google Toolbar. Jedini problem je u tome štiti je navedeni PageRank najčešće star par meseci. Takođe postoje i on-line službe koje omogućuju pregled PageRank-a određene stranice.

Uz PageRank, kod redosleda stranica u obzir se uzimaju i dalji faktori, kao što je zastupljenost ključnih reči unutar teksta ili naslova dokumenata. Takođe veliku ulogu imaju i imena sidra „named anchor“ unutar dokumenata. Poslednja tehnika se veoma često zloupotrebljava kod izrade takozvane „Google bombe“. Tačan naziv Google-ovih rezultata za pretraživače bavi se navedenim metodama sa ciljem da postigne što bolji rezultat uz što manje troškove za vlasnika stranice.

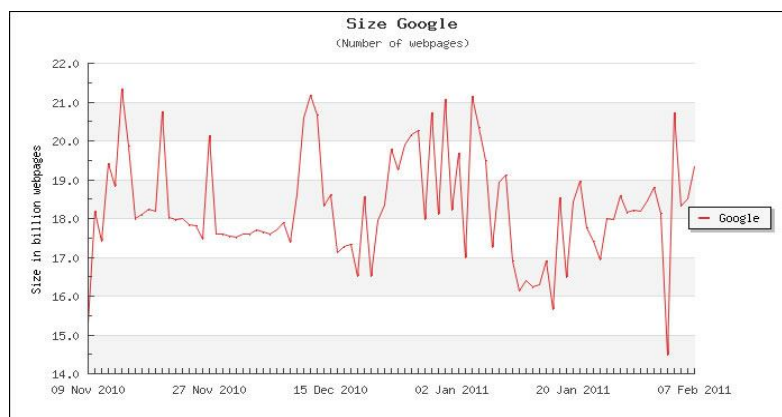
11.4 Veličina index-a

Okviran broj dokumenata u indeksu pretraživača za vreme prvih sedam godina:

- Januar 1998: 25.000.000
- August 2000: 1.060.000.000
- Januar 2002: 3.083.000.000
- Septembar 2004: 4.285.000.000
- Novembar 2005: 8.058.044.651

Pored ovoga, u Junu 2005. godine prema izjavi Google-a postojao je indeks od 1.187.630.000 fotografija, 1.000.000.000 unsent članaka, 6.600 kataloga i oko 4.500 novinskih izvora. Od sedmog rođendana kompanije, veličina indeksa se više ne prikazuje na glavnoj stranici. Prema određenim stručnjacima detaljno numerisanje stranica unutar Google indeksa ne postoji.

Okvirna veličina indeksa Google-a u poslednja tri meseca:



11.5 Usluga pretraživanja za portale

Google-ova tehnika pretraživača sa licencom se dodeljuje raznim Internet portalima, kako bi se i tamo omogućilo pretraživanje Interneta, bez toga da korisnik napušta portal.

11.6 Google appliance

Google appliance je oblik u kojem Google prodaje vlastitu tehniku pretraživanja kompanijama koje je žele koristiti u intranetu. Kod Google appliance-a se radi o serveru sa predinstaliranim softverom, koji u mreži kompanije preuzima ulogu koju inače Google ima unutar World Wide Web-a.

12. Domaći search engine

Domaći Internet je veoma mali. Globalna mreža u našoj zemlji dostupna je korisnicima od 1998. godine dok je u 1997. godini postojala samo Akademska mreža. Domaći pretraživač predstavlja posebnu vrstu specijalizovanog pretraživača, koji je ograničen na govorno područje. Sve opšte vrste pretraživača, o kojima je bilo reči mogu se naći i kod nas, što znači da i svi problemi koje imaju oni, imaju i domaćim uz još mnogo jezičnih problema.

12.1 Problemi domaćih pretraživača

Možda najveći problem domaćih pretraživača jeste odrediti koji je to domaći sajt. Da li su to svi sajtovi koji su na .rs domenu, postoji veliki broj sajtova koji su na engleskom jeziku a nalaze se na .rs domenu. Oni nikako ne spadaju u domaće sajtove. Skoro polovina naših sajtova se nalazi na .com domenu. Njih treba indeksirati. Tu su i ostali domen .net, .org, .gov... Dolazimo do zaključka da su domaći sajtovi oni koji su pisani za naše govorno područje. Problem je u tome, kako napisati algoritam za kroler koji pravi razliku između srpskog i bosanskog koji su veoma slični jezici. Neophodno je pomenuti da smo jedna od retkih zemalja koje imaju dva pisma. ćirilica koja je službeni jezik, i latinica koja je mnogo više zastupljena na web-u zbog sličnosti sa alfabetom. To znači da bi naš pretraživač morao da nam za zahtev na latinici da rezultate pretrage i na ćirilici i obrnuto, što je zaista komplikovana procedura. Međutim ni to nije sve što se pisama tiče, domaći pretraživač bi morao da nam za zahtev na latinici da i rezultate pretrage za taj isti zahtev na alfabetu, i obrnuto. Sledeća jezička karakteristika našeg govornog područja, je to što mi imamo sedam padeža, pa je poželjno da pretraživač ima sposobnost da prepozna koren reči.

	Krstarica.com	Pogodak.rs	Google.rs	Aladin.com
Latinica	podržava	podržava	podržava	podržava
Ćirilica	podržava	podržava	podržava	podržava
Latin+alfabet	ne podržava	podržava	podržava	podržava
Oba pisma	ne podržava	podržava	podržava	ne podržava
Padeži	ne podržava	podržava	ne podržava	ne podržava

Trenutno najpopularniji domaći, kao i svetski pretraživač jeste Google. Srpska verzija Google-a ne podržava dvojezične rezultate pretrage i veoma često kao i drugi pretraživači izlistava sajtove na hrvatskom, bosanskom pa čak i ruskom jeziku, što kod www.pogodak.rs nije sluča

13. SEO - Search engine optimization

SEO (search engine optimization) jeste proces poboljšanja količine i/ili kvaliteta saobraćaja ka stranici (web page) ili sajtu (web site) koji stiže preko pretraživača (search engine). Ove tehnike podrazumevaju izmene HTML koda, dizajna i organizacije stranice u cilju da stranice postanu „search engine friendly“. SEO je u suptornosti sa SEM-om (search engine marketing). Prvi adresira prirodne („natural“, „organic“, „un-paid“, „algorithmic“) pretrage, a poslednji radi sa sponzorisanim ili plaćenim pretragama.

The screenshot shows a Google search interface with the query 'automobili' entered in the search bar. The search bar includes a 'Search' button and a 'Instant is on' dropdown. Below the search bar, it indicates 'About 19,400,000 results (0.09 seconds)' and a link to 'Advanced search'. The results are divided into two main sections: 'Ads' (sponsored results) and organic search results. The first organic result is highlighted with a yellow background and is titled 'Birate nov automobil?'. It includes a brief description and a link to 'www.opelkomnenovic.rs'. The second organic result is titled 'Dug sit i pare na broju' and includes a link to 'www.keskrediti.rs'. The third organic result is titled 'Firestone Auto Repairs' and includes a link to 'FirestoneCompleteAutoCare.com'. The fourth organic result is titled 'Polovni automobili - auto oglasi, prodaja automobila, vozila, auto ...' and includes a link to 'www.polovniautomobili.com/'. The fifth organic result is titled 'Auto oglasi - Polovni automobili - Polovnjaci - Polovna vozila' and includes a link to 'www.kupiauto.com/'. The sixth organic result is titled 'Mojauto - Auto oglasi, polovni automobili, vesti, testovi' and includes a link to 'www.mojauto.rs/'. The seventh organic result is titled 'Automobili Lamborghini Holding Spa' and includes a link to 'www.lamborghini.com/'. The 'Ads' section on the right includes two results: 'Mazda6 od € 17.990' and 'Nađi Mercedes Benz online'.

automobili X Search Instant is on ▾

About 19,400,000 results (0.09 seconds) Advanced search

► **Birate nov automobil?** Ads
Komnenović ima rešenje za Vas. Najbolja ponuda OPEL automobila.
www.opelkomnenovic.rs

Dug sit i pare na broju
Povoljnije refinansiranje 0% kamate na dozvoljeni minus
www.keskrediti.rs

Firestone Auto Repairs
Visit Firestone For All of Your Auto Repair and Service Needs.
FirestoneCompleteAutoCare.com

Polovni automobili - auto oglasi, prodaja automobila, vozila, auto ... 🔍
- [Translate this page]
Najposećeniji sajt za polovne automobile u Srbiji. Kupite ili prodajte auto putem besplatnih oglasa. Pronađite pravi auto za sebe.
Pretraga polovnih vozila - Motocikli - Delovi i oprema - Marke
www.polovniautomobili.com/ - Cached - Similar

Auto oglasi - Polovni automobili - Polovnjaci - Polovna vozila 🔍 - [Translate this page]
Auto oglasi, Polovni automobili Srbija, polovnjaci sa fotografijama.Kupujete ili prodajete automobili, oglasi polovna vozila.
www.kupiauto.com/ - Cached - Similar

Mojauto - Auto oglasi, polovni automobili, vesti, testovi 🔍 - [Translate this page]
Vodeći portal za prodaju automobila u Srbiji. Polovni automobili, novi automobili, besplatni auto oglasi iz cele Srbije. Vesti iz sveta automobila.
www.mojauto.rs/ - Cached - Similar

Automobili Lamborghini Holding Spa 🔍
Official website of Lamborghini's US distributor. Details and photos of every model made since the company's inception, some classifieds.
www.lamborghini.com/ - Cached - Similar

Ads
Mazda6 od € 17.990
Pitajte lokalno prodajno mesto za zoom-zoom ponude-saznajte više!
www.mazda-rs.com

Nađi Mercedes Benz online
Među 21000 polovnih automobila nalazi se i Mercedes kakvog tražiš
www.polovniautomobili.com

See your ad here »

Na Google-u pretraga izgleda kao na slici, gde su prva dva rezultati markirani žutom bojom i rezultati sa desne strane sponzorski i plaćeni linkovi, a sve ostalo su rezultati organske pretrage.

SEO i SEM su od značaja, jer visoko kotiranje u pretrazi garantuje posete ka stranici. Posle toga je na samoj stranici da posetioce zadovolji, zadrži ili konvertuje u kupce (određenih proizvoda ili usluga).

13.1 Web crawl-eri

Kako je za SEO vrlo značajan odnos sa pretraživačima, treba proučiti i način rada pretraživača. Iz tog razloga u radu će osim samih SEO tehnika biti opisano i kako pretraživači sakupljaju informacije o stranicama (indeksiranje) kao i kako treba dizajnirati i pisati HTML kod stranice. SE (search engine), kao što su Google, Msn, Yahoo, koriste automatizovan proces prikupljanja informacija o stranicama (indeksiranje). Za ovo se koristi takozvani web crawler-i (Googlebot, Msnbot, Yahoo Slurp...). S obzirom da je broj Internet stranica ogroman, da se stranice često menjaju i da ima stranica koje su generički stvorene, crawler-i neće obići svaku stranicu i koristiće određene podatke sa same stranice i sa drugih stranica za odluke u procesu crawl-ovanja.

Web crawling politike:

- Selection - koje stranice preuzeti
- Re-visit - kada posetiti ponovo stranice
- Politeness - izbeći overloading stranice
- Parallelization - koordinisati distribuirane web crawler-e

13.1.1 Selection politika

Potrebna je metrika za utvrđivanje relevantnosti stranice. Kao što su unutrašnji kvalitet stranice, popularnost po posetama, linkovima pa čak i popularnost samog URL-a. Različiti pristupi su:

- Restricting followed links – preuzimaju se najčešće samo HTML resursi, ispituje se da li se URL završava sa .html, .htm, .asp, .aspx, .php, ili sa /. Takođe se izbegavaju stranice koje sadrže '?' (dinamička stranica) zbog mogućih spider trap-ova.
- Path-ascending crawling – prikupljanje što više resursa sa stranice na osnovu putanje. Na primer <http://llama.org/hamster/monkey/page.html>, crawl-ovaće se /hamster/monkey/, /hamster/, i /. Ovo je obično u cilju sakupljanja resursa (za galeriju slika recimo kao što je google images) pa se ovakvi crawl-eri nazivaju i harvester softver.
- Focused crawling – relevantnost stranice se utvrđuje sličnošću stranice sa upitom korišćenim u pretrazi. Većina ovih informacija se doprema pre posete stranice kroz anchor text, sam URL stranice i meta podataka. Ovi crawl-eri se nazivaju i topical crawler-i.
- Crawling the Deep Web – većina stranica na Internetu se nalaze u takozvanoj dubokoj mreži (deep web) i nevidljivi su za web crawl-ere jer im se može pristupiti samo upitom. Zato, ako nema linka ka toj stranici neće ni biti posećena. Mapa stranice može omogućiti pretragu ovih stranica (Google's Sitemap Protocol i mod oai ovo omogućavaju). Googlebot može crawl-ovati i sav tekst u hipertekstu: sadržaj, tekst i/ili tag.

URL stranica treba da ima što logičniji naziv koji odgovara sadržaju stranice i zahtevu korisnika. Putanja do resursa na stranici treba da pruži što više informacija, a ne da bude generička. Treba koristiti meta podatke i sav tekst treba da bude smislen: nazivi resursa, putanje do njih, hiperlinkovi, tagovi. Treba formirati sajt mapu – naravno, ne samo zbog crawlera već i zbog korisnika.

13.1.2 Re-visit politika

Stranice na Internetu se menjaju sa vremenom, a zbog njihovog velikog broja proces crawl-ovanja može trajati nedeljama pa i mesecima. Mnoge stranice mogu biti poprilično zastarele dok se proces završi. Uvedeni su pojmovi svežine (freshness) i starosti (age) stranice. Stranica je sveža ako je lokalna kopija ista sa stranicom na mreži, a starost stranice odgovara vremenu proteklom od poslednje izmene lokalne kopije stranice. U prvom slučaju treba smanjiti broj zastarelih stranica, a u drugom treba obezbediti da vreme starosti bude što manje.

Postoje dva pristupa:

- Uniformni - sve stranice su iste i ponovo se posećuju istom učestanošću
- Proporcionalni - češća poseta stranicama koje se menjaju češće

Optimalno rešenje je uniformnom pristupu, ali kada se stranice koje se često menjaju izbace iz stranica za posetu.

13.1.3 Politeness politika

Postojanje velikog broja crawl-era i njihova sama priroda (ogromno korišćenje resursa mreže, istovremeno preuzimanje velikih stranica...) može značajno opteretiti servere. Delimično rešenje problema je robots exclusion protocol poznatiji kao robots.txt protokol. Ovaj fajl govori crawlerima kojim delovima web servisa ne treba da pristupaju. Ovaj protokol ne definiše predlog intervala poseta, iako je to najefikasniji način odbrane od preopterećenja servera. Aks Jeeves, MSN i Yahoo su uveli dodatni parametar „crawl-delay“ u robots.txt. Prvo predloženo rešenje od 60s je bilo neodrživo. Crawler-i uglavnom koriste 1-15s delay između zahteva kao apsolutnu vrednost ili 10t kao relativnu vrednost (t je vreme potrošeno da se preuzme prethodna stranica izraženo u sekundama).

13.1.4 Parallelization politika

Paralelni crawler je crawler koji pokreće više procesa u paraleli. Cilj je maksimalno povećati brzinu preuzimanja uz minimizovanje overhead-a od paralelizacije i izbegavanje višestrukog preuzimanja

13.2 SEO principi i tehnike

Kako je značaj pretraživača rastao tako su se razvijale i SEO tehnike. Nakon što je postao jasan marketinški značaj visokog kotiranja na pretraživačima, pojavile su se i kompanije koje pružaju SEO usluge, pa sam akronim SEO može stajati i za search engine optimizators. Naravno, ubrzo je došlo na zloupotreba poznavanja rada crawler-a, te se SEO tehnike mogu

podeliti na bele (white hat) ili legalne i crne (black hat) ili nelegalne poznatije kao Spamdexing.

Indeksirati se – tj. biti posećen od strane crawl-era. Neki pretraživači (Yahoo!) garantuju indeksiranje plaćanjem. Međutim niko ne garantuje i poziciju u pretrazi. Postoje i direktorijumi koji manuelno pregledaju sadržaj stranice, klasifikuju ga i odlučuju o indeksiranju: Yahoo! Directory i Open Directory Project (ODP). Google recimo nudi Google Webmaster Tools za test stranice i pravljenje sajtmape i robots.txt-a.

Crawl-eri koriste i druge faktore kada odlučuju hoće li indeksirati stranicu ili ne. Jedan od njih može biti i udaljenost stranice od polazne (home) stranice ili direktorijuma (root). Sprečiti neželjeno crawl-ovanje – crawler-i traže u root direktorijumu robots.txt kojeg parsiraju. U tom fajlu je moguće navesti koje stranice crawl-eri ne treba da obilaze. Primer takvih stranica su login specifične stranice kao što su korpe i user-specifičan sadržaj. Google je upozorio i da treba sprečiti indeksiranje internih rezultata pretrage, jer će biti smatrane spam-om.

Povećati relevantnost – pomoću podizanja važnosti stranice. Ugrubo se može podeliti na:

- broj linkova ka stranici: - linkovi ka stranici u okviru jednog sajta (PageRank) i spojašnji linkovi (Cross linking). Treba dodati da spoljni linkovi ka stranici (sa drugih sajtova) povećavaju verovatnoću indeksiranja, ali samo ako su same stranice sa kojih dolaze linkovi relevantne. *Black hat* tehnike su link farming i comment spam.
- optimizacija za upite u pretrazi – korišćenje ključnih reči (*keywords*) na stranici. Najpoznatija *Black hat* tehnika je keyword stuffing.
- optimizacijom URL stranice - URL normalization

13.2.1 White hat i black hat

Bele tehnike su dobro poznate i prihvatljive i sami pretraživači upoznaju korisnike Interneta sa njima u cilju što relevantnije i kvalitetnije pretrage Interneta. S druge strane, crne tehnike su nepoželjne i sankcionišu se pa se treba upoznati i sa njima. Neke kompanije koje pružaju SEO usluge su zbog agresivnih i crnih tehnika kažnjene izbacivanjem sa liste indeksiranih stranica i negativnom reputacijom, što je naravno pogodilo sve njihove klijente, jer su i njihove stranice bile izbrisane sa listi za indeksiranje, a njihovo ime i proizvodi su stavljeni na liste za ignorisanje. Neke tehnike nisu ni bele ni crne i spadaju u takozvane sive (gray hat) tehnike. Primer je kupovina linkova (sa nekog sajta, koji je dobro rangiran, kupi se link da pokazuje na željeni sajt). Google je protiv kupovine linkova, ali mnogi se pretplaćuju na magazine, članstvo ili druge resurse s ciljem da dobiju link ka svojoj stranici. Još jedna siva tehnika je pravljenje mikro-sajtova sa ciljem da sadrže linkove ka željenom sajtu/stranici. Ovo nije „dozvoljeno“, jer je vlasnik stranice isti, ali to pretraživači ne mogu da znaju. Sive tehnike donose određeni rizik, bele tehnike daju dugoročne rezultate dok crne tehnike dovode do privremenog ili trajnog sankcionisanja stranice. Crne tehnike uglavnom služe za online prevare – vek im je kratak (dok ih pretraživači ne otkriju) pa treba da generišu profit za to kratko vreme. Sive tehnike mogu upotrebljavati razni SEO provajderi te treba biti oprezan u izboru SEO partnera.

13.2.2 Spamindexing - black hat SEO

Content spam – izmena sadržaja stranice sa ciljem izvrtnja slike koju pretraživači stiču o stranici.

Keyword stuffing – popunjavanje stranice prekomernim brojem ključnih reči. Ključne reči mogu biti i nevezane za sadržaj same strane. Današnji pretraživači mogu da analiziraju stranicu i otkriju keyword stuffing. Većina pretraživača će trank-ovati velike stranice.

Hidden (invisible) unrelated text – tekst „nulte“ veličine, iste boje kao i pozadina, tekst u „no-frame“ delu...iako ne mora biti spamindexing, (već pristupačnost – accessibility) pretraživači mogu privremeno ili trajno blokirati sajt koji ima nevidljivi tekst na nekoj sebi.

Meta tag stuffing – ponavljanje ključnih reči u meta tagovima i postavljanje meta ključnih reči nevezanih za sadržaj stranice. Ova tehnika je neefektivna od 2005.

Gateway or doorway pages – stranica sa vrlo malo sadržaja za korisnika, ali krcata sličnim ključnim rečima i frazama (dizajnirana da se visoko kotira u pretrazi). Uopšteno, ovakve stranice će imati „click here to enter“ na stranici.

Scraper sites (Made for AdSense) – sajtovi/stranice nastali „scrape“-ovanjem rezultata pretrage pretraživača. Sadržaj im je jedinstven, ali sačinjen od informacija sa drugih stranica (često bez dozvole). Uglavnom su prepuni oglasa (naročito pay-per-click) ili upuću korisnike ka nekoj drugoj stranici/sajtu.

Article spinning – za razliku od scraper sites tehnike, sadržaj i članci sa drugih stranica se „prepakuju“ i ponovo pišu da bi se izbegli penali pretraživača usled dupliranog sadržaja. Proces može biti ručan (pisci) ili automatizovan preko neuralni mreža ili thesaurus baze podataka.

Link spam – korišćenje osobina link-baziranih ranking algoritama. Koriste se visoko rangirane stranice da sadrže linkove ka drugim stranicama, jer time povećavaju njihov rang (linkovi imaju „težinu“) ili se lažira popularnost linkova veštačkim umrežavanjem stranica.

Link farms – grupa gusto povezanih stranica koje se međusobno referenciraju.

Hidden links – skrivanje linkova tako da ih posetioci ne vide s ciljem povećanja popularnosti linka. Posećeni linkovi podižu rang.

„Sybil“ attack – lažno predstavljanje. Spamer formira više sajtova u različitim domenima koji su međusobno povezani. Primer su lažni blogovi (fake/spam blogs).

Spam blogs (splogs)– blogovi koji služe samo za spamovanje, po prirodni slični link farm-ama.

Page hijacking – pravljenje kopija popularnih sajtova, sličnog izgleda i sadržaja kao i originala, koji preusmerava korisnike ka nerelevantnim ili zlonamernim sadržajima.

Buying expired domains – neki spameri prate DNS zapise za domene koji će uskoro isteći, zatim ih kupuju i zamenjuju stranice sa linkovima ka njihovim stranicama. Google resetuje podatke o isteklim domenima pa ova strategija nije korisna za Googlebot.

Cookie stuffing – postavljanje affiliate kolačića bez znanja posetioca sajta. Kolačić će generisati zaradu za osobu koja vrši cookie stuffing. Ovo ne samo da generiše lažirane affiliate prodaje već ima potencijal da „pregazi“ ostale, legitimne, affiliate kolačiće i ukrade im profit.

User word-writable page – sajtovi koji omogućavaju da posetioci vrše izmene na njima, ako nisu dobro anti-spam zaštićeni, mogu biti iskorisćeni za ubacivanje linkova ka spam sajtovima.

Spam in blogs - postavljanje linkova i ključnih reči na slučajan način, nevezano za sadržaj samog sajta. Guest books, forumi, blogovi i bilo koji sajt koji dozvolja komentare posetilaca podložan je ovom napadu.

Comment spam – slično spamovanju blogova. Samo se odnosi na bilo koji sajt koji dozvoljava izmene, kao što je recimo vikipedija. Vrlo je problematično zbog mogućnosti automatizovanog napada.

Wiki spam – specijalan slučaj comment spama, karakterističan za vikipediju. 2005. vikipedija je implementirala podrazumevanu „nofollow“ politiku za „rel“ HTML atribut pa se likovi ignorišu od strane Google-ovog PageRank algoritma.

Refereer log spamming – neki sajtovi koriste referrer log koji pokazuje koje stranice su povezane sa tim sajtom (referrer je stranica sa koje je neka druga stranica referisana). Pisanjem robot-a (programa za automatski pristup, obilaženje i delanje na Internetu) koji po random principu posećuje veliki broj sajtova mnogo puta, sa porukom ili specifičnom adresom datoj od referrer-a, ta poruka ili adresa se pojavljuje u referrer logovima. Kako neki pretraživači koriste ovaj log za utvrđivanje relevantnosti stranice/sajtova, referrer log se može koristiti za spam.

Mirror websites – hostovanje više sličnih sajtova korišćenjem različitih URL-a.

URL redirection – preusmeravanja korisnika ka drugim sajtovima bez njegove intervencije, na primer korišćenjem meta refresh-a, flash-a, java skriptova, java...

Cloacking – korišćenje tehnike ili tehnika u kojima se prikaz posetiocu i crawl-eru razlikuje. Klokovanje može biti korišćeno i za povećanje pristupačnosti (accessibility) sajta osobama sa posebnim potrebama ili za obezbeđivanje sadržaja koje pretraživači ne bi mogli da obrade. Sam Google koristi vrstu klokovanja (IP delivery) za dostavljanje rezultata.

Code swapping je zlonamerna upotreba klokovanja – stranica se optimizuje za top rang, a po visokom kotiranju se zamenjuje drugom stranicom.

13.2.3 SEO kao marketinška strategija

Visoko rangiranje je važno i to su potvrdila istraživanja po kojima najveću verovatnoću za posetu imaju rezultati pretrage koji se nalaze u gornjem levom uglu rezultata. Međutim, SEO je samo jedan od načina da privučete korisnike na stranicu, sam sadržaj stranice je ono što je bitnije za zadržavanje korisnika i prodavanje proizvoda i usluga sa stranice. Ipak se sve radi zbog korisnika. Veoma je bitno reći i da se sami proces indeksiranja i rangiranja pretrage menja, često bez najave, i sam mehanizam indeksiranja je tajna. To dalje znači da je izrazito nesigurno osloniti se mnogo na SEO kao glavni marketinški stub. SEO je bitan, ali ne treba da bude i jedini način marketinga. Pokazalo se da linkovi sa određenih stranica mogu daleko više doprineti prodaji usluga i proizvoda nego sam SEO.

13.3 Google search engine friendly stranica

U ovom poglavlju su sažete Google smernice za dizajn stranice koja će biti search engine friendly. Uopšteno govoreći, sav tekst treba da je jedinstven, precizan, jasan i relevantan u odnosu na sadržaj stranice. Pretraživači kažnjavaju dupliran, veoma sličan ili generički sadržaj i tekst nezvan za sadržaj stranice.

Smernica za dizajn stranice, koje se opisuju u nastavku teksta, su:

- jedinstveni i precizni naslovi
- upotreba meta tagova
- URL struktura
- laka navigacija sajtom
- QoS
- bolji anchor text
- pravilna upotreba heading tagova

- optimizacija upotrebe slika
- efektivna upotreba robots.txt-a
- „nofollow“ linkovi
- promocija sajta
- upotreba besplatnih webmaster alata
- upotreba analitičkih alata i servisa
- resursi za webmaster-e

Jedinstveni i precizni naslovi - title tag govori posetiocima i pretraživačima šta je tema stranice. Idealno, svaka stranica bi trebalo da ima svoj naslov. Sadržaj title tag-a će najverovatnije biti prikazan na prvoj liniji rezultata pretrage. Anatomiju rezultate pretrage možete upoznati preko klipa: Anatomy of search result.

- precizno opisati sadržaj stranice, izbegavati generičke nazive („united“) i nazive nevezane za sadržaj stranice. Ključne reči (iz zahteva pretrage) će biti boldovane.
- jedinstveni naslovi svih stranica, izbegavati isti naslov za sve ili veliku grupu stranica.
- kratki ali opisni naslovi, izbegavati vrlo dugačke naslove i keyword stuffing naslova

Upotreba meta tagova - Dok su naslovi (title tag) koncizni, meta tagovi su više opisni i mogu biti i dužine manjeg paragrafa. Ove informacije se mogu koristiti kao snipete u rezultatima pretrage. Informacije prikazane u pretrazi mogu poticati i sa samog sajta, linka ka sajtu ili ODP-a (Open Directory Project). Ključne reči (iz upita pretrage) će biti boldovane.

- Precizno sumirati sadržaj stranice – da se posetioci informišu i zainteresuju, izbeći nevezanost sa sadržajem, generečke („this is a wabpage“, „page about...“), keyword stuffing, copy-paste celog sadržaja stranice u meta tag.
- Jedinstven opis za svaku stranicu (preciznije određuju koja stranica korisniku treba shodno upitu), izbeći korišćenje istog meta taga za sve stranice ili velike grupe stranica.

URL struktura - Smisleni nazivi stranica, putanja, resursa.

Umesto www.primera.com/a/4567/nesto.htm bolje je www.auto.com/audi/c4/motor.html. URL treba da bude lako razumljiv i čak navigabilan, jer se takav URL lakše pamti. Ako se koriste alati koji proizvode dinamičke stranice (kao Joomla!) na ovo treba posebno obratiti pažnju.

- Koristiti (smislene) reči koje su relevantne sadržaju stranice, izbegavati dugačke URL-e sa nepotrebnim parametrima (session ID...), generička imena, keyword stuffing
- Jendostavna struktura direktorijuma – laka za organizovanje i pruža dobar osećaj posetiocu gde se na sajtu nalazi i koji tip sadržaja je na stranici (slike, video, tekst...), izbegavati mnogo poddirektorijuma (/dir1/dir2/dir3/.../dir6/page.html), upotrebu imena direktorijuma koja nije vezana za sadržaj.
- Jedan URL za pristup dokumentu – ovim se postiže da se reputacija linkova ne rasipa na dva ili više mesta. Ukoliko posetioci dospevaju do istog sadržaja preko različitih linkova u okviru sajta treba setovati 301 redirect sa nepreferiranih URL-a ka dominantnom URL-u . Izbeći pristup istom sadržaju iz root direktorijuma i poddirektorijumima, mešanje www i ne-www URL-a, neumerenu upotrebu velikih slova (većina korisnika očekuje mala slova i bolje ih pamti).

Laka navigacija sajtom - Posetioci treba da se brzo snađu na sajtu i pronađu šta traže. Dobro je imati mapu sajta.

- Prirodna hijerarhija – prirodno i lako kretanje od opšteg ka specifičnom, izbeći složene mreže navigacionih linkova (povezivanje svake stranice sa svakom) i preterano usitnjavanje sadržaja (posetiocu treba mnogo klikova da dođe do željenog)
- Pre svega upotrebljavati tekst za navigaciju, pretraživači i posetioci lakše shvataju i pamte tekstualne linkove. Neki uređaji ne podržavaju Flash ili JavaScript. Izbeći navigaciju zasnovanu u potpunosti na padajućim menijima, slikama ili animacijama.
- Putanja navigacije (breadcrumbs): start>prva>druga>cilj
- HTML sajtmapa i XML sitemap fajl – prvi za korisnika, drugi za pretraživače. Izbeći neažurnost sajtmape i sajtmape koje nisu organizovana već su samo lista stranica na sajtu.
- Navigacija brisanjem dela URL-a – www.auto.com/audi/c4/motor.html. Šta se desi ako se obriše poslednji deo (ostane www.auto.com/audi/c4/)?
- Korisna 404 stranica – ljubazno uputiti posetioca na drugu (ispravnu, najčešće root) stranicu uz opis mogućih razloga greške. Google-ov alat 404 widget. Izbeći indeksiranje 404 stranice (osigurati 404 HTTP status code), poruke tipa „not found“, „404“ ili nedostatak stranice, korišćenje različitog dizajna (i duha) 404 stranice u odnosu na dizajn sajta. Ponekad je šaljiv pristup dobar:

QoS - Kvalitetan sadržaj je ono što je najbitnije, jer to će zadržati posetioca na sajtu.

- Lako čitljiv tekst – dobro formatiran tekst bez grešaka i sa razumnom upotrebom paragrafa („cigla“ teksta na celoj stranici će lako odbiti posetioce). Treba izbegavati tekst u okviru slike, jer će posetioci možda hteti da prekopiraju taj tekst.
- Ne lutati od teme – lepo organizovati teme uz jasnu predstavu gde se koja tema završava, a druga počinje i u koju temu korisnik trenutno čita. Izbeći veliku količinu teksta različite tematike koji je neorganizovan u paragrafe, podpoglavlja ili odvojen separatorima.
- Upotreba relevantnog jezika – koje su to reči koje će korisnik tražiti (između ostalog i u upitu na pretraživaču)? Ko je korisnik i koje će reči razumeti (da li je ekspert za oblast ili ne)?
- Svež i jedinstven sadržaj – novi sadržaj će zadržati stare korisnike i privući nove. Izbeći prepakivanje ili kopiranje postojećeg sadržaja sa malo dodatne vrednosti za posetioce
- Ekskluzivni sadržaj ili servis – nov, korisni sadržaj koji niko drugi ne nudi...istraživanja, vesti, iskorišćenje baze korisnika...drugi sajtovi mogu imati manjak iskustva ili resursa za tako nešto.
- Sadržaj je primarno za posetioce, ne za pretraživače, te treba izbegavati nepotrebne ključne reči, zlonamerno skrivati tekst od korisnika ali prikazivati ga pretraživačima, blokove teksta koje imaju malu vrednost posetiocima ("frequent misspellings used to reach this page").

Bolji anchor text - Tekst koji će posetioći videti kao rezultat linka (piše se u anchor tagu), pretraživačima govori o stranici koja je linkovana sa stranicom sa koje je link.

- Opisan tekst, a ne generički (page, article, klik here...), nerelevantan tekst, URL linkovane stranice
- Konciznan tekst – kratko i jasno, par reči ili kratka fraza
- Lako uočivi linkovi – jasno se razlikuju od običnog teksta, korisnik ne treba da ih promaši, ali ni da ih slučajno klikne. Uobičajeno je da su neposećeni linkovi plave boje. Izbeći upotrebu CSS koja pravi linkove tako da se ne razlikuju od ostatka teksta
- Sve ovo važi i za interne linkove, pomaže korisnicima u navigaciji, ali i šretraživačima. Izbeći preteranu upotrebu ključnih reči, dugačak tekst (samo zbog SEO-a) i nepotrebne linkove koji ne doprinose navigaciji na sajtu.

Pravilna upotreba heading tagova

- Ne treba mešati sa header ili head tagom. Postoji 6 heading tagova (<h1> do <h6>) poređanih po opadajućoj veličini slova (i samim tim važnošću teksta). Njihova upotreba doprinosi boljoj hijerarhiji, organizovanju i lakšem shvatanju sadržaja stranice.
- Outline – koje su najbitnije stvari, šta treba istaći, kako sve poređati po važnosti? Izbeći haotično skakanje sa jednog heading taga na drugi, neadekvatnu upotrebu (možda je bolji), postavljanje nekorisnog teksta koji remeti strukturu stranice
- Kvalitetno upotrebljavati heading tagove - upotreba treba da je smisljena, previše heading tagova može otežati čitanje i snalaženje na stranici. Izbeći preteranu upotrebu, stavljanje celog teksta stranice u heading tag, upotreba samo za stilizovanje umesto za struktuiranje teksta

Optimizacija upotrebe slika - Smisljena imena i alt atribut treba iskoristiti. Alt tekst je vrlo bitan za SEO, jer se tretira slično anchor tekstu. Slike se takođe mogu pretraživati (Google images).

- Kratka i opisna imena i alt tekstovi, izbeći generička i dugačka imena i keyword stuffing
- Alt tekst pri upotrebi slike kao linka – da i pretraživač zna o čemu je reč. Izbeći veoma duge alt tekstove i upotrebu samo slika kao linkova na sajtu.
- Smeštanje slika u zasebnom direktorijumu za njih – lakša organizacija i jednostavnije putanje
- Upotreba uobičajenih tipova – JPEG; GIF; PNG; BMP.

Efektivna upotreba robots.txt-a - Fajl za pretraživače smešten u root direktorijumu. Služi da kaže pretraživačima koje stranice ne treba da indeksira. Google-ov alat je robots.txt generator

- Sigurne metode za osetljiv sadržaj – ne treba se osloniti samo na robots.txt, jer se sama URL može referencirati i neki crawler-i ne koriste robots protokol. Dakle, enkripcija i zaštita lozinkom je preporučena. Izbeći indeksiranje rezult stranica (posetioći mrže da sa jedne pretrage dođu na drugu pretragu), velikog broja autogenerisanih stranica sličnog sadržaja i stranica koje su rezultat proxy servisa .

„Nofollow“ linkovi - Postavljanje rel atributa na „nofollow“ će reći pretraživačima da ne prati dotični link na stranici ili da ne predaje reputaciju stranice linkovanoj stranici. Ovo je korisno za blogove sa javnim komentarima – sprečavaju spam in blogs. Isto važi za bilo koji deo sajta kojeg posetioци mogu menjati.

- Blogovanje novih sadržaja i servisa – obavestiti korisnike o novom sadržaju
- Offline promocija – URL sajta na vizitkama, lećima, pismima, čestitkama...
- Social media sajtovi – sajtovi koji su organizovani oko određene grupe ljudi s ciljem njihovog povezivanja sa određenim sadržajem ili servisima. Izbegavati promociju svake sitnice – treba pucati visoko. Izbeci veštačku promociju sadržaja
- Dodavanje na Google's Local Business Center - Local Business Center omogućava kontakt sa potencijalnim kupcima preko Google Maps-a i pretrage.
- Povezivanje sa srodnić sajtovića – Nove ideje, projekti i saradnja. Mogućnost komplementarne saradnje (obguskator + licensiranje + enkripcija). Izbeci link spaming i kupovinu linkova sa drugih sajtova.

Upotreba besplatnih webmaster alata - Na Internet ima mnoštvo besplatnih alata za proveru i optimizaciju sajtova.

Upotreba analitićkih alata i servisa - Ove alate koristimo kako bi saznali: kakav je saobraćaj na sajtu, preko kojih zahteva za pretragu posetioци stižu na sajt, sa kojih sajtova stižu, kako se ponašaju na samom sajtu, sta je najposećenije, da li je neki sadržaj izrazito nepopularan. Neki od alata su:

- Google Analytics
- Google Website Optimizer
- Google Webmaster Tools

toga je potrebno optimizaciju izvoditi pažljivo, a nepotvrđene akcije ne izvoditi odjednom u nizu, već jednu po jednu, kako bi mogli pratiti efekte SEO.

14.2 Odabir ključnih reči

Izbor ključnih reči bi trebao da bude prvi korak pri konstruisanju sajta. Trebalo bi da imate listu ključnih reči dostupnih za unošenje u tekst pre nego što počnete sa izgradnjom sajta. Da bi definisali ključne reči, trebalo bi pogledati SEO servise kao što su www.wordtracker.com i slični, koji su dobro početno mesto. Ali treba znati da podatke koje nude se ponekad mogu razlikovati od onih koji bi najviše odgovarali vašem sajtu. Takođe, treba da znate da Google ne daje informacije o frekvenciji upita pretraživanja. Nakon približnog određivanja ključnih reči, možete da analizirate ključne reči sajtova Vaše konkurencije i pokušate da odredite koje ključne reči će te koristiti.

Postoje dve jasne strategije odabira ključnih reči - optimizovati sajt za mali broj visoko popularnih ključnih reči ili optimizovati sajt za veliki broj manje popularnih ključnih reči. U praksi se obe strategije često kombinuju.

Nedostak ključnih reči koje su veoma popularne je to što je njihova vrednost kod popularnih sajtova mnogo veća nego kod novih sajtova. Često je nemoguće da se novi sajtovi prikažu na vrhu pretrage u pretraživačima za ove pretraživačke upite.

Za ključne reči koje su mnogo ređe, upiti u pretraživačima je često dovoljno samo pomenuti potrebne kombinacije reči na stranici ili uraditi minimalnu optimizaciju teksta. U određeni okolnostima ove reči vam mogu doneti veliki broj poseta.

Kada završite sa listom ključnih reči, trebalo bi da odredite glavnu ključnu reč za koju ćete optimizovati Vaš sajt. Retke reči nisu dobar izbor. One će se najverovatnije svakako nalaziti na sajtu. Kao pravilo, vrednost većine popularnih fraza je veoma visoka kod konkurencije. Zbog toga morate da sagledate jasnu sliku vaše konkurencije. Da bi ste dostigli vrednost konkurencije morate da ocenite broj parametara za prvih 10 sajtova koji se pojave kao rezultat pretrage:

1. Prosečan PageRank tih sajtova
2. Prosečan broj linkova ka tim sajtovima
3. Broj stranica na Internetu sa istom ciljnom grupom, i ukupan broj rezultata.
4. Broj stranica na Internetu koje imaju iste fraze ključnih reči.

Ovi podaci će vam pokazati koliko je teško da Vaš sajt dostigne na samom vrhu za te ključne reči. Nakon ovoga možete odabrati nekoliko srednje popularnih ključnih reči sa prihvatljivom vrednošću kod konkurencije, koje možete koristiti da promovišete i optimizujete sajt.

Kao što je već spomenuto, podaci koje nude razni servisi za SEO, kao i sami pretraživači, često znaju da budu pogrešni. To znaci da se jako retko mogu podesiti optimalne ključne reči iz prvog pokušaja. Nakon nekog vremena ćete najverovatnije moći da vidite koje su to ključne reči koje bi mogle da se dodaju vašem sajtu.

Ukoliko otkrijete loše ključne reči, možda bi trebalo da potražite neke bolje. Ukoliko vidite da ključne reči koje ste odabrali nisu optimizovale vaš sajt kako ste želeli. Ako se opredelite za ovaj potez morate uneti novi prerađeni set ključnih reči. Što znači da ce te morati da rekonstruišete sajt promenom sadržaja teksta tako što ce te dodati nove ključne reči i dodavanjem novih stranica koje sadrže nove ključne reči.

```

1 <!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">
2 <html xmlns="http://www.w3.org/1999/xhtml">
3 <head>
4 <meta name="google-site-verification" content="jGOoV6RYG8GHQSCE8V8Ccia3tFqk3Zq_NU8aBpFGOrA" />
5 <META name="y_key" content="b5cfe8e2cc28d535" />
6 <meta name="description" content="ELZING Elektronska zastita inzenjering">
7 <meta name="robots" content="all">
8 <meta name="keywords"
9 content="Elzing, Elektronska zastita inzenjering, Video nadzor, Alarmi, Alarmni sistemi, Kontrola pristupa, Intrefoni, Kamere, Ip Kamere, Analogne kamere, bezbednost, zastita, Graficki dizajn,
10 Web dizajn, Satelitski i kablovski sistemi, Evidencija radnog vremena, Vizuelni identitet, Sajtovi, vizit karte, www.elzing.rs, procena ugrozenosti objekta, procena rizika i mere zastite,
11 pustanje sistema u rad ">
12 <LINK REL="SHORTCUT ICON" HREF="http://www.elzing.rs/favicon.ico">
13 <meta http-equiv="Content-Type" content="text/html; charset=utf-8" />
14 <script type="text/javascript">
15
16 var _gaq = _gaq || [];
17 _gaq.push(['_setAccount', 'UA-19214968-1']);
18 _gaq.push(['_trackPageview']);
19
20 (function() {
21 var ga = document.createElement('script'); ga.type = 'text/javascript'; ga.async = true;
22 ga.src = ('https:' == document.location.protocol ? 'https://ssl' : 'http://www') + '.google-analytics.com/ga.js';
23 var s = document.getElementsByTagName('script')[0]; s.parentNode.insertBefore(ga, s);
24 })();
25
26 </script>
27 <title>Početna stranica</title>
28 <style type="text/css">
29 <!--
30 #apDiv1 {
31 position:relative;
32 width:1005px;
33 height:220px;
34 z-index:1;
35 background-color: #FFFFFF;
36 margin-top: 10px;
37 margin-right: auto;
38 margin-bottom: 0px;
39 margin-left: auto;
40 }
41 #apDiv2 {
42 position:relative;

```

content="Elzing, Elektronska zastita inzenjering, Video nadzor, Alarmi, Alarmni sistemi, Kontrola pristupa, Intrefoni, Kamere, Ip Kamere, Analogne kamere, bezbednost, zastita, Graficki dizajn, Web dizajn, Satelitski i kablovski sistemi, Evidencija radnog vremena, Vizuelni identitet, Sajtovi, vizit karte, www.elzing.rs, procena ugrozenosti objekta, procena rizika i mere

14.3 Description tag

Meta description ima veliki značaj kada je optimizacija sajta u pitanju. On predstavlja kratki opis sajta, tematiku sajta na određenom jeziku i meta description govori Google i internet korisnicima koja je tema njihovog sajta. Meta description se nalazi u head tagu između <head> i </head>, odmah ispod naslova i pre ključnih reči:

```

<head>
<meta name="google-site-verification"
content="jGOoV6RYG8GHQSCE8V8Ccia3tFqk3Zq_NU8aBpFGOrA" />
<META name="y_key" content="b5cfe8e2cc28d535" />
<meta name="description" content="ELZING Elektronska zastita inzenjering">
<meta name="robots" content="all">
<meta name="keywords"
content="Elzing, Elektronska zastita inzenjering, Video nadzor, Alarmi, Alarmni sistemi,
Kontrola pristupa, Intrefoni, Kamere, Ip Kamere, Analogne kamere, bezbednost, zastita,
Graficki dizajn, Web dizajn, Satelitski i kablovski sistemi, Evidencija radnog vremena, Vizuelni
identitet, Sajtovi, vizit karte, www.elzing.rs, procena ugrozenosti objekta, procena rizika i mere
zastite, pustanje sistema u rad ">
<LINK REL="SHORTCUT ICON" HREF="http://www.elzing.rs/favicon.ico">
<meta http-equiv="Content-Type" content="text/html; charset=utf-8" />
<script type="text/javascript">

```

Meta description tag, bas kao i svi ostali tagovi, nisu vidljivi korisniku i nisu vidljivi na samoj strani, međutim, oni imaju uticaja na drugi način kada je SEO optimizacija sajta za pretraživače u pitanju i to što nisu vidljivi ne umanjuje njihov značaj. Opis sajta koji upišete u ovaj meta tag će se pojaviti u rezultatima pretrage za određene ključne reči, odmah ispod naslova koji je linkovan na Vaš sajt. Na taj način, Internet korisnici koji pretražuju za određene ključne reči, kada pročitaju opis sajta u Google rezultatima pretrage, kliknu na određeni link, ukoliko je taj link ono što njih zanima.

Iz tog razloga je bitno da meta description tag bude jasan sažetak Vašeg sajta, da ne bude predugačak ili prekratak, da ukratko u jednoj ili dve rečenice opiše ono što posetioци mogu očekivati na Vašem sajtu. Opis dakle treba da bude informativan, koncizan, ne suviše prenatrpan ključnim rečima i da ukratko opiše ono što korisnik može da pronađe na tom sajtu.

Google preporučuje da se za svaku stranu na jednom sajtu napravi unikatan meta description, da se tu smesti ono što nije moglo da se opiše u naslovu sajta, da bude čitljiv i razumljiv. kvalitetan i da ukratko predstavi sadržaj te strane. Dužina meta description taga treba da bude ne veća od 150-170 karaktera ili jedna do dve rečenice. Google webmaster tools pruža veoma korisnu analizu sadržaja i između ostalog pruža informacije o meta description tagovima na određenom sajtu, ukoliko su suviše kratki, predugi ili ukoliko na sajtu imate duplirane naslove i opise. Description meta tags su važni iz razloga što ih Google možda može koristiti prilikom prikazivanja sajta u svojim rezultatima pretrage

Upotreba robots.txt fajla i robots meta tagova

Robots.txt fajl je običan tekstualni fajl pomoću koga je moguće dati određene instrukcije pretraživačima ili web robotima u pogledu toga koje stranice na sajtu da indeksiraju, a koje ne. Na primer, Google ima svoje web robote tj. programe koji pretražuju internet i indeksiraju sve web stranice na koje naiđu. Pomoću robots.txt fajla moguće je reći Googlu da neku stranicu ili ceo direktorijum ne indeksira i samim tim ne prikazuje u rezultatima pretraživanja. Dakle, pomoću robots.txt fajla se govori botovima pretraživača šta da indeksiraju na sajtu a šta ne, tako da je pravljenje ovog fajla jedna od prvih stvari koje se rade u procesu optimizacije sajta za pretraživače.

Da bi se napravio robots.txt fajl moguće je koristiti običan Notepad. Primer jednog osnovnog robots fajla bio bi:

```
User-agent: *  
Disallow: /
```

Sa ove dve linije zabranjuje se svim pretraživačima da indeksiraju bilo šta sa web sajta. Naravno, ovo je pod pretpostavkom da pretraživač koji je posetio sajt poštuje Robots Exclusion Protocol, što glavni pretraživači i čine (za razliku od spam ili malware botova).

Još neke naredbe koje se mogu napisati u robots.txt fajlu:

```
User-agent: Yahoo! Slurp  
Disallow: /neki-folder
```

Ako bi vam iz nekog razloga trebalo da baš Yahoo web crawler koji je poznat kao Slurp ne indeksira neki folder, to biste postigli sa gornjom naredbom. Znači, pomoću naredbe Disallow: / može se sakriti od pretraživača neki folder ili neki fajl. Spisak svih Web crawlera ili botova možete pogledati na ovoj adresi <http://www.user-agents.org/>

User-agent: Google
Disallow:
User-agent: *
Disallow: /

na ovaj način Google robot ima pristup celom sajtu (samo se izostavi kosa crta / posle Disallow:) dok je svim ostalim robotima zabranjeno da pristupe.

User-agent: *
Disallow: /cgi-bin/
Disallow: /tmp/
Disallow: /junk/

na ovaj način se brani pristup samo navedenim direktorijumima, dok je ostatak sajta slobodan za indeksiranje.

Robots.txt može biti korisan u raznim situacijama, na primer dok je sajt još uvek u fazi izrade ili testiranja, ako postoje dve verzije istog sajta ili istog fajla, ako ne želite da se neka stranica nađe u rezultatima pretraživanja itd.

Osim posebnog robots.txt fajla koji se postavlja u root direktorijum sajta na web serveru (www.sajt.com/robots.txt), mogu se koristiti i robots meta tagovi, koji se ubacuju u head deo html dokumenta. Moguće varijante su:

```
<META NAME="ROBOTS" CONTENT="NOINDEX, FOLLOW">  
<META NAME="ROBOTS" CONTENT="INDEX, NOFOLLOW">  
<META NAME="ROBOTS" CONTENT="NOINDEX, NOFOLLOW">
```

Prilikom korišćenja robots.txt fajla treba imati u vidu da pojedini roboti mogu jednostavno da ignorišu robots.txt fajl (malware ili spam botovi), kao i da je robots.txt fajl moguće lako videti. Drugim rečima, ne koristite robots.txt fajl da biste sakrili važne informacije na sajtu.

14.4 XML Mapa sajta

XML mapa sajta je spisak svih Web-stranica na sajtu. Namenjena je isključivo robotima pretraživača, a ne posetiocima sajta koji je, po pravilu, neće ni videti.

XML mapa sajta informiše robote pretraživača o svakoj strani na Web-sajtu, uključujući URL adrese ili strane koje pretraživači možda ne bi sami otkrili, kao što su, na primer, dinamične strane. Jednostavno rečeno, Sitemap XML osigurava da Google i drugi pretraživači saznaju za svaku stranu na vašem Web-sajtu.

XML mapa sajta je datoteka koja omogućava webmasterima da uključe dodatne informacije o svakom URL-u, kao na primer:

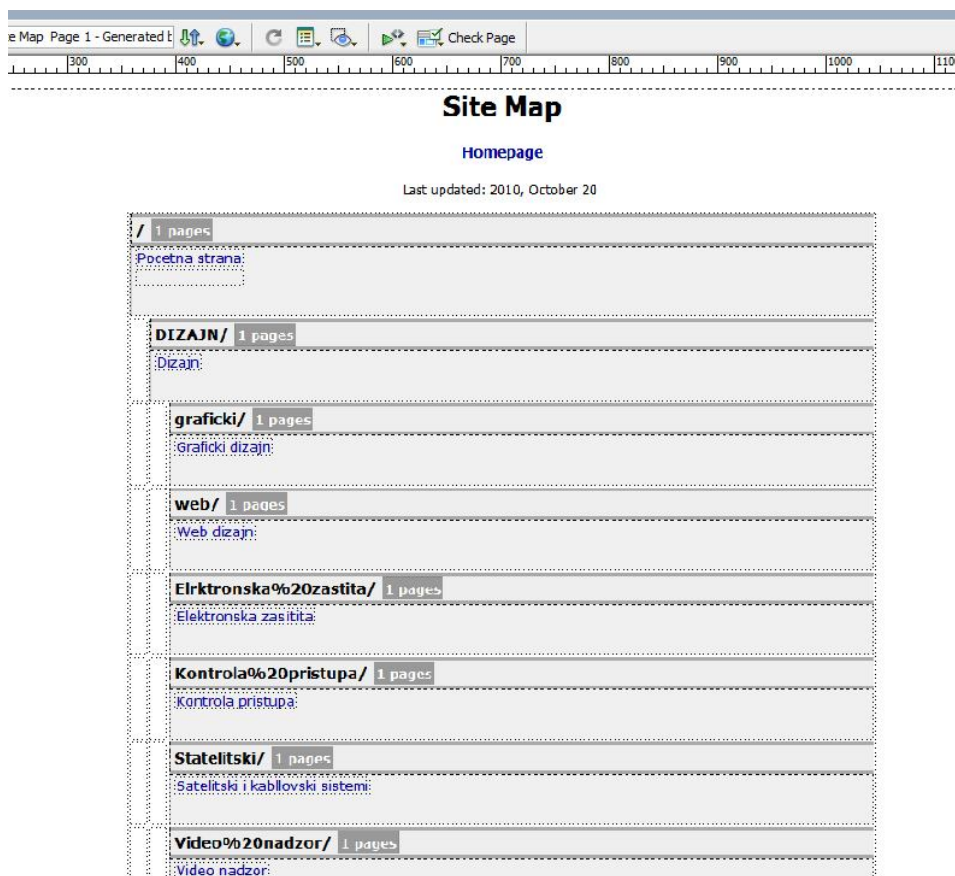
- datum kad je URL adresa poslednji put ažurirana, to jest, menjana;
- koliko često se Web-strane na sajtu menjaju;
- Prioritet ili relativnu važnost stranice u odnosu na druge strane na sajtu. Prioritet koji se dodeli određenoj strani ne utiče na njenu poziciju u rezultatima pretraživanja. Pretraživači koriste ove informacije kad biraju između strana u okviru istog sajta, tako da možete da iskoristite ovu mogućnost da povećate verovatnoću da se važnije strane na vašem sajtu nađu u indeksu pretraživača. Pošto je prioritet ili važnost stranica

relativna; važnost strana na vašem sajtu neće da se upoređuje sa važnošću strana na drugim sajtovima.

Kreiranje i prijavljivanje XML mape sajta pretraživačima se posebno preporučuje ako:

1. Vaš Web-sajt sadrži dinamičan sadržaj.
2. Web-sajt sadrži strane koje roboti ne mogu tako lako otkriti. Na primer, strane koje koriste AJAX, Flash ili slike.
3. Web-sajt je nov i mali broj linkova vodi ka njemu sa drugih Web-sajtova. Googlebot (Guglov robot) indeksira veb tako što prati linkove od jedne veb-strane do druge i ako vaš sajt nije dovoljno povezan sa drugim Web-sajtovima, onda će robotu trebati više vremena da ga pronađe.
4. Web-sajt sadrži veliku arhivu stranica sa sadržajem koje nisu uopšte međusobno povezane ili su loše povezane.

XML mapu sajta možete da kreirate pomoću alatki za generisanje mape sajta. Kad kreirate mapu sajta, možete da je pošaljete pretraživačima. Ako želite da pošaljete svoju mapu sajta Google-u, to možete da uradite pomoću Google alatki za vebmastere.



14.5 Važnost linkova na sajtu

Da bi se imala dobra pozicija u pretraživačima, potrebno je imati dosta odlaznih kao i dolaznih linkova od i do sajt. Dolazni linkovi su jako važni, od njih najviše zavisi vaša pozicija na Googlu i ostalim pretraživačima, kao i posećenost sajta. Međutim, ne treba ignorisati ni

važnost odlaznih linkova. Kvalitetni izlazni linkovi sa sajta mogu indirektno poboljšati poziciju u pretraživačima, jer ukazuju da je sajt autoritet u nekoj oblasti. Isto tako, linkovi ka lošim (spam) sajtovima, link direktorijumima ili tematski ne povezanim sajtovima utiču na kredibilitet sajta koji linkuje. Ne postoje garancije da ćete imati dobru poziciju na Googlu samo zato što linkujete ka drugim kvalitetnim i relevantnim sajtovima, ali indirektno SEO stručnjaci se uglavnom slažu da je ovo jedan od faktora na koji treba obratiti pažnju.

ELZING

SURVEILLANCE SECURITY DESIGN WEB

IP Video nadzor

Prelaz na otvorene video sisteme, u kombinaciji sa olakšanjima koje donosi mreža, digitalnim projektovanjem i inteligencijom kamere, stvaraju se mnogo efikasnija sredstva nadgledanja sigurnosti i daljinskog monitoringa nego što je to ranije bilo. Mrežni video donosi sve što analogni video nudi, a pored toga i širok spektar novih funkcija i karakteristika koje omogućuje digitalna tehnologija.

Prilikom izbora odgovarajućeg sistema, korisnik je uporediti tehnologije koje su dostupne u pogledu zahteva područja primene, skalabilnosti, jednostavnosti upotrebe i fleksibilnosti. Mrežni video, koji se inače shvata kao IP video nadzor u okviru sigurnosnog nadzora i daljinskog monitoringa, predstavlja sistem koji korisnicima daje mogućnost praćenja i snimanja video preko IP mreže (LAN/WAN/Internet). Za razliku od analognih video sistema, mrežni video koristi mrežu umesto point-to-point kablova, kao osnovu (backbone) prenosa informacija. U aplikacijama mrežnog video, tokovi digitalnog video se prenose na bilo koje mesto u svetu preko žične ili bežične IP mreže, omogućujući pri tome monitoring i snimanje sa bilo kog dela mreže.



Većina je ispariti kamere koje odgovaraju vašim potrebama. To uključuje kamere koje mogu biti pan/tilt/zoom, vandalproof, vaterproof ili fixne dome. Danas su dostupni svi tipovi mrežnih kamera i bez obzira kakve su vaše potrebe mrežne kamere su u mogućnosti da ih ispunje. Iako je dostupno mnoštvo analognih kamera mrežne kamere sada nude dodatne prednosti uključujući bolji kvalitet slike i više fleksibilnosti prilikom instalacije. A za specijalne instalacije gde je potrebna veoma visoka rezolucija ili bežična mrežna kamera je jedina opcija. Štovanje pravne mrežne kamere je od suštinske važnosti za uspeh sistema. Međutim nisu sve mrežne kamere iste. Ako koristite jeftine mrežne kamere moguće je da kvalitet slike ne bude dovoljno dobar. Mnogo je stavki koje čine kvalitet mrežne kamere.

Pojavom prvih IP kamera, koje daju digitalni video signal u potpunosti je zadovoljen tehnološki nivo analognih kamera. Razvojem digitalne mikro elektronike i usvajanjem najnovijih standarda dolazi do naglog porasta tehničkih karakteristika i











14.6 Prijava sajta na pretraživače

Kada se prijavljuje URL na pretraživače, vi zapravo govorite pretraživaču da imate novu Web stranu, i pozivate ga da i tu stranu uključi u rezultat pretrage. Naravno pretraživač će i bez ovog prijavljivanja pronaći vašu stranu ali prijava se radi kako bi se skratio vremenski period. Prijava Web sajta na pretraživač je veoma jednostavan proces.

Add your URL to Google

Share your place on the net with us.

We add and update new sites to our index each time we crawl the web, and we invite you to submit your URL here. We do not add all submitted URLs to our index, and we cannot make any predictions or guarantees about when or if they will appear.

Please enter your full URL, including the `http://` prefix. For example: `http://www.google.com/`. You may also add comments or keywords that describe the content of your page. These are used only for our information and do not affect how your page is indexed or used by Google.

Please note: Only the top-level page from a host is necessary; you do not need to submit each individual page. Our crawler, Googlebot, will be able to find the rest. Google updates its index on a regular basis, so updated or outdated link submissions are not necessary. Dead links will 'fade out' of our index on our next crawl when we update our entire index.

URL:

Comments:

Optional: To help us distinguish between sites submitted by individuals and those automatically entered by software robots, please type the squiggly letters shown here into the box below.



Add URL

<http://www.google.com/addurl/?continue=/addurl>

Sve što je potrebno da se uradi je da na stranici za dodavanje Url adrese u polje URL unese adresa vašeg Web sajta, a u polje Comments možete da dodate komentar ili ključne reči koje opisuju vašu stranicu.

Submit Your Site

Submit a Website or Webpage

Enter the URL for the website or webpage you would like to submit. For any URL (directly submitted or obtained from a feed) our crawler will extract links and find pages we have not discovered already. We will automatically detect updates on pages and remove dead links on an ongoing basis.

Submit URL

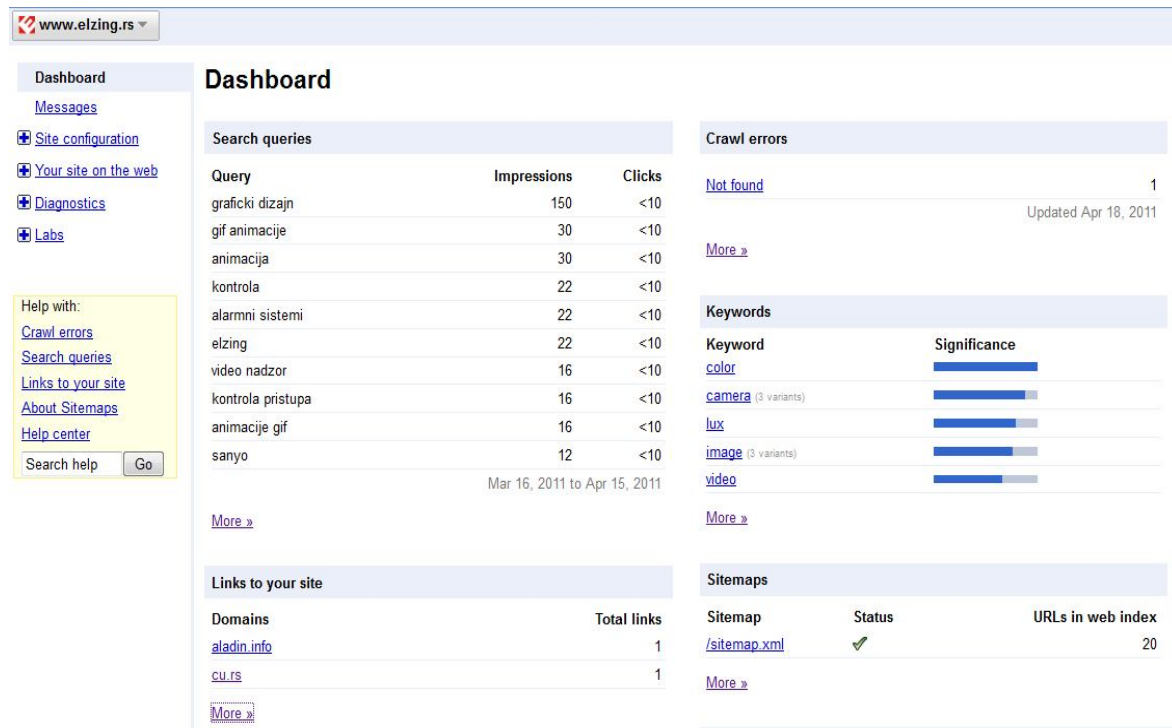
Please include the `http://` prefix (for example, `http://www.yahoo.com`).

Submit Site Feed

<http://siteexplorer.search.yahoo.com/submit>

14.7 Google webmaster tools

Web master tools je veoma korisna alatka preko koje može da se: uploaduje mapa sajta (.xml), da se uradi monitoring sajta, ključnih reči, rankinga, i drugih faktora koji utiču na poziciju Web strane.



14.7.1 Site Configuration

Sitemaps - Dodajte vasu mapu sajta.

Crawler access - Podesite stranice koje ne želite da budu u pretrazi. Recimo forumi imaju na svim stranama link ka registraciji a taj URL vam i nije potreban na svakoj strani sajta pa ga bolje izbacite.

Sitelinks – najčešće se prikazuju stranice koje se nalaze na svim stranicama sajta ili one stranice koje su dobile najviše eksternih linkova sa drugih sajtova. Moguće da je da uklonite one koje ne želite da se pojavljuju u pretrazi.

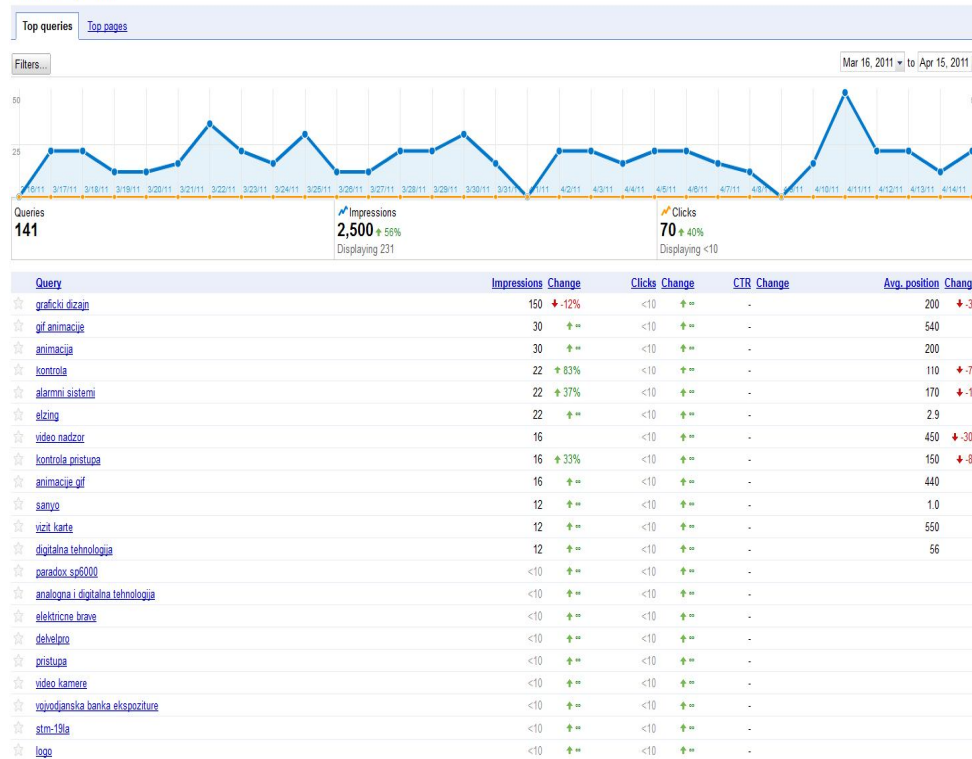
Change of address – prijavite promenu domena. Ima i odlično objašnjenje kako da to uradite pomocu 301 redirekcije

Settings - Ako je Vaš sajt na srpskom , onda odaberite iz padajućeg menija Srbiju. Druge dve opcije su za određivanje učestalosti koliko će često da vas obilazi google i na koji način će biti prikazan Vaš URL.

14.7.2 Your site on the web

Search queries - možda i najbitniji deo Google webmaster centra. Pokazuje pojmove koji su najrelevantniji vašem sadržaju, poziciju na Google-u . Odnos pregledanih strana i klikova (CTR). Mogućnost da označite zvezdicom pojmove koji vas interesuju. Iz padajućeg menija za vaš sajt odaberite Srbiju i opciju WEB.

Search queries



Links to your site - prikazuje linkove ka vašem sajtu i anchor text , odnosno reči koje su linkovane ka web stranici, što je veoma bitno za SEO.

Internal links - koji interni linkovi sa vašeg sajta se najčešće prikazuju i oni koji se najviše koriste.

Keywords - prikazuje ključne reči koje su relevantne za vaš sadržaj i lista istih zavisi od toga koliko se one često prikazuju u vašim tekstovima.

www.elzing.rs

Dashboard

Messages

Site configuration

Your site on the web

Search queries

Links to your site

Keywords

Internal links

Subscriber stats

Diagnostics

Labs

Help with:

My site isn't doing well in search

Keywords

Google-friendly sites

Help center

Search help

Go

Keywords

Below are the most common keywords Google found when crawling your site. These should reflect the subject matter of your site.

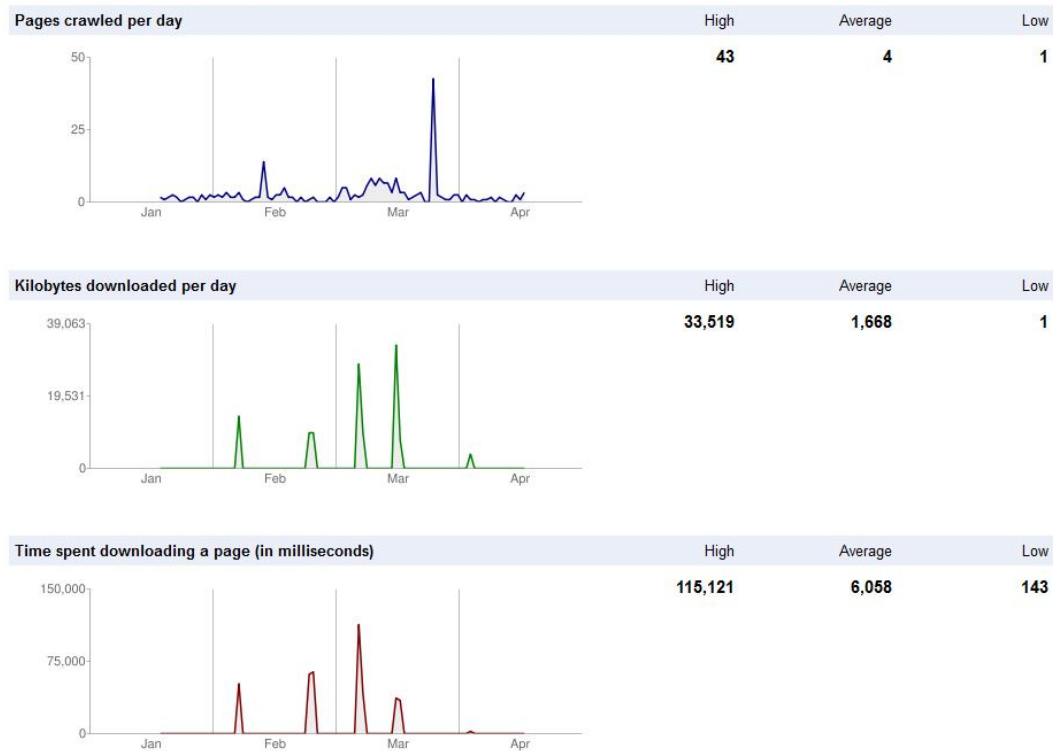
Keyword	Significance
1. color	
2. camera (3 variants)	
3. lux	
4. image (3 variants)	
5. video	
6. resolution	
7. max	
8. control (2 variants)	
9. auto	
10. digital (2 variants)	
11. samsung (2 variants)	
12. min	
13. lines	
14. lens (2 variants)	
15. support (4 variants)	
16. sens	
17. 12v	
18. dimension (2 variants)	
19. manual	
20. night	
21. power (2 variants)	
22. ccd	
23. class	
24. noise	

14.7.3 Diagnostic

Crawl stats - odličan pokazatelj koliko je vašem sajtu potrebno da se učita, koliko troši protoka i koliko se stranica indeksira na dnevnom nivou.

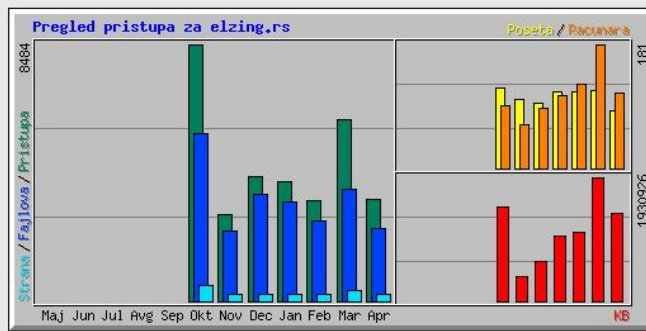
Crawl stats

Googlebot activity in the last 90 days



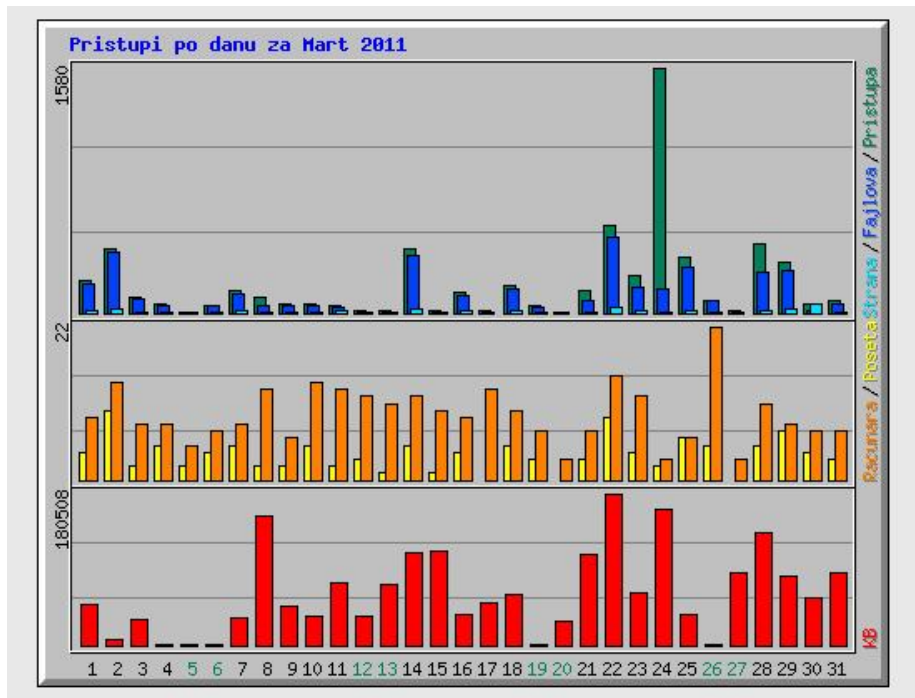
HTML suggestions – Prikazuje sugestije koje treba da izmenite. Dupli nazivi stranica, dupli opisi.

14.8 Statistika za „www.elzing.rs“

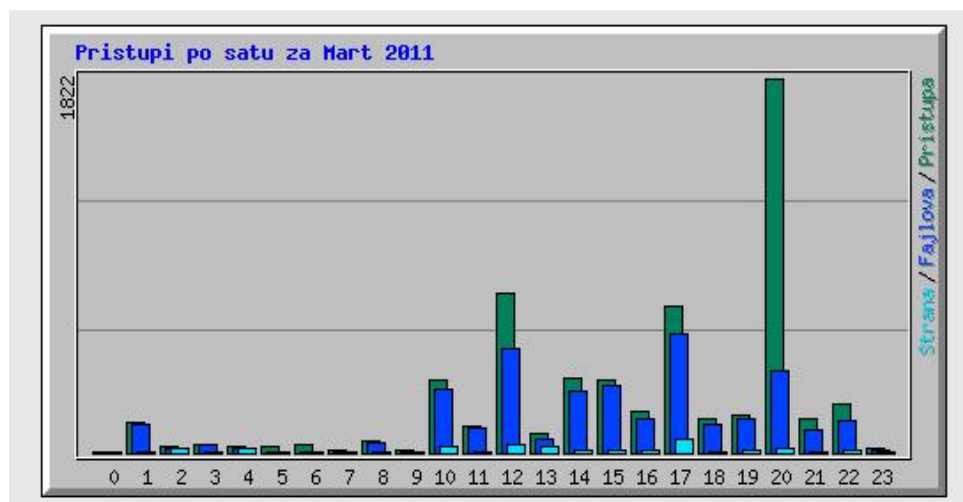


Pregled po mesecima										
Mesec	Dnevni prosek					Mesecni zbrovi				
	Pristupa	Fajlova	Strana	Poseta	Racunara	KB	Poseta	Strana	Fajlova	Pristupa
Apr 2011	186	134	12	4	110	1368862	84	220	2420	3361
Mar 2011	193	118	10	3	181	1930926	113	338	3685	5998
Feb 2011	119	94	8	4	122	1077727	112	244	2656	3334
Jan 2011	127	105	8	3	106	1006734	112	248	3265	3955
Dec 2010	132	113	7	3	88	624464	96	230	3524	4120
Nov 2010	95	77	7	3	63	381039	101	231	2323	2852
Okt 2010	652	426	41	9	92	1468391	118	543	5550	8484
Ukupno za sve						7858143	736	2054	23423	32104

Mesecna statistika za Mart 2011		
Ukupno pristupa	5998	
Ukupno fajlova	3685	
Ukupno Strana	338	
Ukupno Poseta	113	
Ukupno KBajta	1930926	
Ukupno razlicitih racunara	181	
Ukupno razlicitih URL-ova	161	
Ukupno razlicitih referisanja	68	
Ukupno razlicitih browsera	91	
	prosek	Max
Pristupa na sat	8	1482
Pristupa na dan	193	1580
Fajlova na dan	118	491
Fajlova na sat	10	54
Racunara na dan	5	22
Poseta na dan	3	10
KBajta na dan	62288	180508
Pristupa po kodu		
Code 200 - OK	61.44%	3685
Code 206 - Partial Content	27.84%	1670
Code 304 - Not Modified	3.68%	221
Code 404 - Not Found	7.04%	422

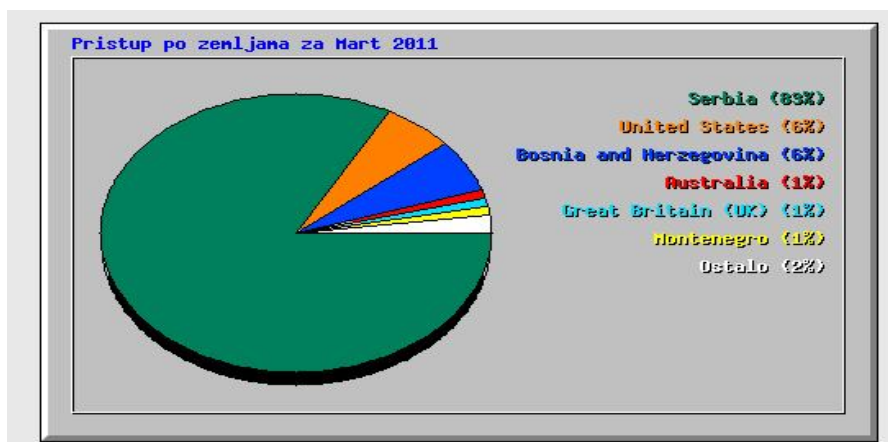


Dnevna statistika za Mart 2011												
Dan	Pristupa		Fajlova		Strana		Poseta		Racunara		KB	
1	210	3.50%	183	4.97%	14	4.14%	4	3.54%	9	4.97%	47759	2.47%
2	411	6.85%	390	10.58%	29	8.58%	10	8.85%	14	7.73%	7472	0.39%
3	97	1.62%	91	2.47%	2	0.59%	2	1.77%	8	4.42%	30590	1.58%
4	54	0.90%	50	1.36%	5	1.48%	5	4.42%	8	4.42%	1626	0.08%
5	6	0.10%	2	0.05%	2	0.59%	2	1.77%	5	2.76%	5	0.00%
6	53	0.88%	48	1.30%	4	1.18%	4	3.54%	7	3.87%	1271	0.07%
7	140	2.33%	120	3.26%	12	3.55%	5	4.42%	8	4.42%	32347	1.68%
8	102	1.70%	53	1.44%	3	0.89%	2	1.77%	13	7.18%	153178	7.93%
9	54	0.90%	45	1.22%	2	0.59%	2	1.77%	6	3.31%	47356	2.45%
10	63	1.05%	51	1.38%	5	1.48%	5	4.42%	14	7.73%	34665	1.80%
11	52	0.87%	41	1.11%	16	4.73%	2	1.77%	13	7.18%	73918	3.83%
12	21	0.35%	9	0.24%	3	0.89%	3	2.65%	12	6.63%	34322	1.78%
13	16	0.27%	8	0.22%	1	0.30%	1	0.88%	11	6.08%	71784	3.72%
14	416	6.94%	367	9.96%	26	7.69%	5	4.42%	12	6.63%	109559	5.67%
15	20	0.33%	6	0.16%	1	0.30%	1	0.88%	10	5.52%	112698	5.84%
16	131	2.18%	116	3.15%	11	3.25%	4	3.54%	9	4.97%	36668	1.90%
17	17	0.28%	6	0.16%	0	0.00%	0	0.00%	13	7.18%	49645	2.57%
18	178	2.97%	159	4.31%	11	3.25%	5	4.42%	10	5.52%	60237	3.12%
19	50	0.83%	40	1.09%	3	0.89%	3	2.65%	7	3.87%	549	0.03%
20	7	0.12%	2	0.05%	0	0.00%	0	0.00%	3	1.66%	28531	1.48%
21	143	2.38%	85	2.31%	6	1.78%	3	2.65%	7	3.87%	108022	5.59%
22	566	9.44%	491	13.32%	37	10.95%	9	7.96%	15	8.29%	180508	9.35%
23	240	4.00%	167	4.53%	19	5.62%	4	3.54%	12	6.63%	61900	3.21%
24	1580	26.34%	152	4.12%	10	2.96%	2	1.77%	3	1.66%	160879	8.33%
25	364	6.07%	297	8.06%	16	4.73%	6	5.31%	6	3.31%	36241	1.88%
26	85	1.42%	80	2.17%	5	1.48%	5	4.42%	22	12.15%	1379	0.07%
27	13	0.22%	5	0.14%	0	0.00%	0	0.00%	3	1.66%	85674	4.44%
28	444	7.40%	264	7.16%	12	3.55%	5	4.42%	11	6.08%	134591	6.97%
29	327	5.45%	278	7.54%	22	6.51%	7	6.19%	8	4.42%	83226	4.31%
30	62	1.03%	19	0.52%	54	15.98%	4	3.54%	7	3.87%	57164	2.96%
31	76	1.27%	60	1.63%	7	2.07%	3	2.65%	7	3.87%	87160	4.51%



Statistika po satu za Mart 2011												
Sat	Pristupa			Fajlova			Strana			KB		
	prosek	Ukupno		prosek	Ukupno		prosek	Ukupno		prosek	Ukupno	
0	0	7	0.12%	0	1	0.03%	0	3	0.89%	461	14285	0.74%
1	4	150	2.50%	4	137	3.72%	0	7	2.07%	1962	60816	3.15%
2	0	27	0.45%	0	19	0.52%	0	18	5.33%	785	24342	1.26%
3	1	43	0.72%	1	40	1.09%	0	2	0.59%	10	297	0.02%
4	1	34	0.57%	0	23	0.62%	0	18	5.33%	2381	73818	3.82%
5	0	29	0.48%	0	4	0.11%	0	1	0.30%	1006	31176	1.61%
6	1	43	0.72%	0	8	0.22%	0	5	1.48%	469	14541	0.75%
7	0	13	0.22%	0	6	0.16%	0	2	0.59%	136	4222	0.22%
8	1	61	1.02%	1	50	1.36%	0	2	0.59%	2393	74177	3.84%
9	0	15	0.25%	0	8	0.22%	0	4	1.18%	1843	57141	2.96%
10	11	355	5.92%	9	309	8.39%	0	27	7.99%	4866	150861	7.81%
11	4	129	2.15%	4	124	3.36%	0	5	1.48%	296	9166	0.47%
12	25	775	12.92%	16	509	13.81%	1	40	11.83%	7699	238674	12.36%
13	3	98	1.63%	2	66	1.79%	1	32	9.47%	3260	101053	5.23%
14	11	363	6.05%	9	305	8.28%	0	14	4.14%	3457	107154	5.55%
15	11	351	5.85%	10	331	8.98%	0	15	4.44%	4110	127411	6.60%
16	6	201	3.35%	5	170	4.61%	0	10	2.96%	3241	100472	5.20%
17	23	713	11.89%	18	580	15.74%	2	68	20.12%	4652	144208	7.47%
18	5	162	2.70%	4	143	3.88%	0	8	2.37%	2423	75122	3.89%
19	5	184	3.07%	5	163	4.42%	0	11	3.25%	1972	61139	3.17%
20	58	1822	30.38%	13	403	10.94%	0	26	7.69%	5659	175432	9.09%
21	5	169	2.82%	3	113	3.07%	0	5	1.48%	2903	89981	4.66%
22	7	234	3.90%	5	161	4.37%	0	13	3.85%	3531	109461	5.67%
23	0	20	0.33%	0	12	0.33%	0	2	0.59%	2773	85976	4.45%

Prvih 30 od 161 URL-ova					
#	Pristupa		KB		URL
1	1409	23.49%	96349	4.99%	/Katalozi/KATALOZI/VIDEO NADZOR/DAHUA KATALOG.pdf
2	125	2.08%	1301466	67.40%	/Katalozi/KATALOZI/VIDEO NADZOR/SAMSUNG KATALOG.pdf
3	92	1.53%	26546	1.37%	/Katalozi/KATALOZI/VIDEO NADZOR/ARECONT KATALOG.pdf
4	75	1.25%	60	0.00%	/favicon.ico
5	73	1.22%	536	0.03%	/Scripts/AC_RunActiveContent.js
6	70	1.17%	1286	0.07%	/SLIKE/animacija za logo.swf
7	68	1.13%	148	0.01%	/SLIKE/dugmici/grafickidizajn2.png
8	68	1.13%	127	0.01%	/SLIKE/dugmici/reference2.png
9	68	1.13%	234	0.01%	/SLIKE/dugmici/satelitski1.png
10	67	1.12%	9375	0.49%	/SLIKE/Gornja animacija.swf
11	67	1.12%	85	0.00%	/SLIKE/dugmici/dizajn1.png
12	67	1.12%	219	0.01%	/SLIKE/dugmici/evidencijaradnogvremena1.png
13	67	1.12%	141	0.01%	/SLIKE/dugmici/grafickidizajn1.png
14	66	1.10%	122	0.01%	/SLIKE/dugmici/alamnisistem1.png
15	66	1.10%	88	0.00%	/SLIKE/dugmici/dizajn2.png
16	66	1.10%	152	0.01%	/SLIKE/dugmici/kontrolapristup1.png
17	66	1.10%	112	0.01%	/SLIKE/dugmici/webdizajn1.png
18	66	1.10%	115	0.01%	/SLIKE/dugmici/webdizajn2.png
19	65	1.08%	119	0.01%	/SLIKE/dugmici/alamnisistemi2.png
20	65	1.08%	203	0.01%	/SLIKE/dugmici/elzastita1.png
21	65	1.08%	208	0.01%	/SLIKE/dugmici/elzastita2.png
22	65	1.08%	233	0.01%	/SLIKE/dugmici/evidencijaradnogvremena2.png
23	65	1.08%	79	0.00%	/SLIKE/dugmici/interfoni1.png
24	65	1.08%	108	0.01%	/SLIKE/dugmici/katalozi1.png
25	65	1.08%	109	0.01%	/SLIKE/dugmici/kontakt2.png
26	65	1.08%	232	0.01%	/SLIKE/dugmici/satelitski2.png
27	65	1.08%	121	0.01%	/SLIKE/dugmici/video nadzor1.png
28	65	1.08%	129	0.01%	/SLIKE/dugmici/video nadzor2.png
29	64	1.07%	80	0.00%	/SLIKE/dugmici/interfoni2.png
30	64	1.07%	109	0.01%	/SLIKE/dugmici/katalozi2.png



15. Web pretraživač

Najefikasniji način pristupa informacijama na Web-u jeste korišćenjem Web pretraživača. U opštem slučaju, pretraživač se sastoji iz tri dela :

- crawler , koji je zadužen za automatsko prikupljanje stranica sa Web-a i njihovo smeštanje u indeks pretraživača
- indexer , koji obezbeđuje kreiranje odgovarajuće strukture (inverted index), koja omogućava efikasnu reprezentaciju i pretraživanje arhiviranih stranica.
- query handler, koji prihvata korisničke upite i odgovara na njih korišćenjem indeksa pretraživača

15.1 Crawling

Web crawling predstavlja proces prikupljanja stranica sa Web-a, radi njihovog indeksiranja u okviru Web pretraživača. Cilj crawling-a jeste prikupljanje što većeg broja Web stranica, zajedno sa informacijama o njihovoj međusobnoj povezanosti, u što kraćem vremenskom periodu i na najefikasniji mogući način.

Funkcionalnosti koje crawler mora da obezbedi :

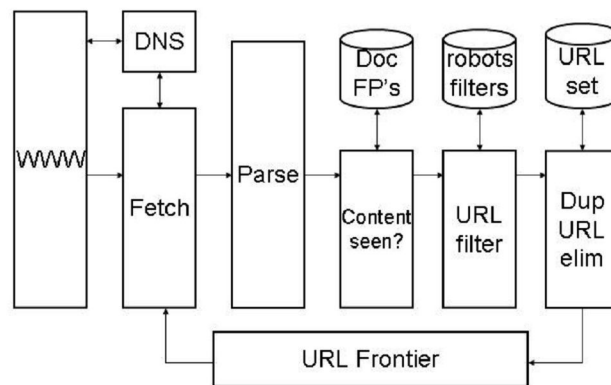
- Robustnost – Na Web-u, postoje serveri koji kreiraju tzv. „spider traps“, zlonamerno generisane web stranice koje imaju za cilj onemogućavanje rada crawler-a, tako što nastoje da kreiraju beskonačnu petlju unutar domena u kojoj se crawler može „zaglaviti“. Crawler-i moraju biti dizajnirani tako da budu otporni na ovaj vid „zamki“ (posebno iz razloga što mogu postojati i greške ovoga tipa koje se ne javljaju namerno, već usled grešaka)
- Pristojnost – Web serveri imaju implicitne i eksplicitne polise po pitanju rate-a kojim ih crawler može posećivati. Ove polise se moraju poštovati.

Funkcionalnosti koje bi crawler trebalo da obezbedi :

- Distribuiranost – Crawler bi trebao da bude u stanju da se izvršava distribuirano na više računara
- Skalabilnost – Arhitektura crawler-a bi trebalo da podržava povećavanje opsega crawl-rate-a dodavanjem dodatnih računara i bandwidth-a
- Performanse i efikasnost – Crawl sistem bi trebao da maksimalno koristi različite sistemske resurse, kao što su procesor, storage i mrežni bandwidth
- Kvalitet – Crawler bi trebao da bude orijentisan ka fetch-ovanju prvo “korisnijih” stranica
- Ažurnost rezultata – Crawler bi trebao da funkcioniše neprekidno, ponovno fetch-ujući sveže kopije prethodno fetch-ovanih stranica. U principu, crawler bi trebao da crawl-uje stranicu sa frekvencijom koja aproksimira frekvenciju promene date stranice
- Proširivost – crawler bi trebao da bude proširiv u više aspekata : da se snađe sa novim formatima podataka, novim fetch protokolima i drugo, što zahteva modularnu arhitekturu crawler-a.

Osnovna operacija svakog hypertext crawler-a je sledeća : Crawler počinje skupom od jednog ili više URL-ova (seed set), odabira URL iz ovog skupa i fetch-uje stranicu na tom URL-u. Nakon toga, parsira fetch-ovanu stranicu, da bi ekstrahovao tekst i linkove iz stranice. Tekst se predaje indeksu, dok se url-ovi dodaju u URL frontier (predstavlja strukturu zasnovanu na redovima za čekanje), koji se u svakom trenutku sastoji od URL-ova čije odgovarajuće stranice trebaju biti fetch-ovane od strane crawler-a. Ovaj postupak se ekvivalentno može posmatrati i kao obilazak Web Grafa. Sam crawler se sastoji od više modula :

1. URL frontier, koji sadrži URL-ove koji će biti fetch-ovani u tekućem crawl-u
2. DNS resolution modul koji određuje adresu web servera na kome se nalazi URL koji fetch-ujemo (ovaj modul se može nalaziti i van modula, tj. može se koristiti sistemski servis za DNS rezoluciju)
3. fetch modul – koji retrieve-uje stranicu na datom URL-u
4. parsing modul – koji ekstrahuje skup linkova sa zadate web strane
5. modul koji određuje da li se ekstrahovani link već u URL frontier redu ili je nedavno fetch-ovan



Crawling se obično obavlja u vidu više thread-ova, od kojih svaki obavlja logički ciklus prikazan na slici 1 . Ovi thread-ovi se mogu obavljati u okviru jednog procesa ili biti partitionisani na više procesa koji se izvršavaju na distribuiranom sistemu.

Crawler thread počinje tako što uzima URL iz frontier-a i fetch-uje web stranu na zadanom url-u, (najčešće koristeći http protokol). Fetch-ovana stranica se smešta na pomoćni storage, gde se obavljaju operacije nad njom. Kao prvo, testira se da li stranica sa istim sadržajem već ne postoji na nekom drugom url (eliminacija duplikata). Dalje, stranica se parsira i ekstrahuju se tekst i linkovi. Tekst se prosleđuje indeksu, kao i informacije o linkovima koji se koriste kao dodatna informacija za rankiranje. Dodatno, svaki od linkova prolazi niz testova da bi se utvrdilo da li ide u URL frontier. Kao prvo, konsultuje se URL filter, u okviru koga se može definisati isključenje određenih domena (npr. svih .com url-ova – vrši crawl svih url-ova osim onih koji pripadaju .com domenu), takođe selekcija može biti i inverzna (crawluju se samo .com url-ovi).

15.1.1 Eliminacija duplikata stranica

Web sadrži višestruke kopije mnogih dokumenata. Po nekim ocenama, 40% svih stranica na Web-u predstavljaju višestruke kopije, istog sadržaja. Pretraživači se trude da izbegnu indeksiranje duplikata, čime ostvaruju uštedu u zauzeću storage kapaciteta, kao i ubrzanje crawl procesa.

Najjednostavniji pristup detekciji duplikata je određivanje fingerprint-a svake web strane, koji predstavlja digest (hash proizvoljne dužine, npr. 64. bita) sadržaja date strane. Na ovaj način, ukoliko dve stranice imaju identične fingerprint-e, stranice se proglašavaju za duplikate. Međutim, ovaj jednostavan pristup nije efikasan u detekciji približnih duplikata (stranice gotovo identičnog sadržaja, sa izuzetkom par karaktera, npr. datum, ime host-a...), kakvi se često javljaju na Web-u.

15.1.2 Case Study: Nutch crawler

Nutch projekat (<http://lucene.apache.org/nutch/>), predstavlja skup open source komponenti, razvijenih na jeziku Java, sa ciljem implementacije celokupnog web pretraživača, na platformi otvorenog koda.

Crawler (Nutch crawler), razvijen u okviru projekta, pored osnovnih funkcionalnosti, podrazumeva i modularnu arhitekturu, koja omogućava proširivanje samog crawler-a, razvojem plugin-ova (na primer za parsiranje i fetch nestandardnih formata, preprocesiranje i slično..)

Osnovu crawler sistema, predstavlja crawl aplikacija i skup konfiguracionih fajlova u okviru kojih se definiše ponašanje i rad samog crawl procesa.

Osnovna konfiguracija definiše se u okviru fajlova nutch-default.xml (koji definiše globalne parametre vezane za crawl uopšte) i nutch-site.xml, u okviru koga se definišu osnovni parametri vezani za konkretnu implementaciju crawler-a (parametri koji se pojavljuju u zaglavlju upita koje šalje crawler – ime crawler-a, opis, url, kontakt email i sl.). Takođe, od posebnog interesa je fajl crawl-urlfilter.txt, u okviru koga možemo, korišćenjem regularnih izraza definisati strukturu url-ova čiji sadržaj prihvatamo ili odbijamo.

Sam crawler se može izuzetno detaljno konfigurisati, u okviru nutch-default.xml fajla. Neke od najznačajnijih konfiguracionih opcija su :

- http.content.limit – maksimalna veličina sadržaja http stranice koja se fetch-uje
- http.redirect.max – maksimalan broj redirekcija koje crawler prati da bi fetch-ovao stranicu
- fetcher.server.delay – dužina pauze koju crawler pravi između dva uzastopna zahteva ka istom host-u
- fetcher.threads.fetch – broj fetcher thread-ova koji se izvršavaju
- fetcher.threads.per.host.by.ip – binarna vrednost koja određuje da li se host-ovi broje po ip adresi ili po hostname-u (tj. da li se podrazumevaju virtuelni ili stvarni host-ovi)

15.2 Indeksiranje

Invertovani indeks, predstavlja osnovnu strukturu podataka koja se korsiti u okviru Web pretraživača i information retrieval softvera uopšte. Pod invertovanim indeksom, podrazumevamo indeks struktru koja sadrži presikavanja između ključnih reči i njihovih lokacija u skupu dokumenata, i korišćenjem koje se omogućava efikasno pretraživanje posmatranog skupa.

Postoje dve osnovne varijante realizacije invertovanih indeksa : na nivou zapisa (record level inverted index), koji se još naziva inverted file index i koji sadrži listu referenci na dokument za svaku reč koja se u okviru njega javlja makar jedanput i na nivou reči (word level inverted index) koji dodatno u okviru indeksa, sadrži i informacije o poziciji svakog javljanja date reči u okviru odgovarajućeg dokumenta.

15.2.1 Metode estimacije veličine indeksa

Iako je relativno jednostavno odrediti fizičku veličinu samog indeksa (u smislu broja dokumenata koji je indeksiran), u slučaju Web pretraživača, tako definisana veličina je od malog značaja, budući da ne pruža informaciju o količini informacija koje se nalaze u indeksu. U praksi je od interesa procenjivanje veličine segmenta Web-a koji indeksira jedan pretraživač. Ovo se pokazuje kao veoma teško za određivanje, budući da u principu postoji praktično beskonačan broj dinamičkih web stranica. (npr. "soft 404 errors" stranice koje Web server automatski generiše). Prema tome, pitanje određivanja veličine indeksa pretraživača, možemo jedino da reformulišemo kao pitanje određivanja relativne veličine indeksa dva zadata pretraživača. Ipak i određivanje ove veličine se takođe pokazuje kao teško, iz sledećih razloga :

1. Pretraživači mogu da vrte web strane čiji sadržaj nisu (potpuno ili čak ni parcijalno) indeksirali. U principu, pretraživači indeksiraju samo prvih par hiljada reči na web strani. U nekim slučajevima, pretraživač je svestan stranice na koju linkuju stranice koje on indeksira, ali sama stranica nije indeksirana. Ovo ipak omogućava da pretraživač vrati smislene rezultate vezane za p u search results

2. Pretraživači u principu, organizuju indekse u vidu više particija, od kojih se ne razmatraju sve u svakom pretraživanju. Na primer, web strana, duboko u okviru web sajta može biti indeksirana ali se ne razmatra u slučaju generalnog web search-a, dok se razmatra u slučaju specifičnih upita vezanih za posmatrani web sajt

Prema tome, pretraživači mogu sadržati višestruke klase indeksiranih strana, tako da ne postoji jedinstvena mera njihovih celokupnih indeksa. Kao odgovor na navedene probleme, razvijen je veliki broj tehnika koje omogućavaju procenu odnosa veličina indeksa dva pretraživača, i . Osnovna hipoteza od koje se polazi jeste da svaki pretraživač indeksira onaj segment Web-a, koji se odabira nezavisno i slučajno sa uniformnom raspodelom. Uvedene pretpostavke : da postoji konačna veličina web-a iz koje pretraživač odabira podskup i da pretraživač odabira iz slučajno odabranog podskupa – su daleko od realnosti. Ipak, navedene pretpostavke predstavljaju dobru osnovu i omogućuju primenu klasične tehnike estimacije poznate kao capture-recapture method

15.2.2 Case Study: Lucene indexer

Lucene projekat (<http://lucene.apache.org/>), predstavlja open source information retrieval biblioteku, implementiranu na jeziku Java, koja omogućuje izuzetno efikasno indeksiranje i pretraživanje, uz mogućnost integracije sa drugim aplikacijama. U okviru Lucene projekta, definisan je API koji omogućava jednostavno kreiranje i ažuriranje invertovanog indeksa. Da bi se dokument dodao u invertovani indeks, prethodno se vrši njegovo skeniranje i dobija se lista podataka, koji opisuju frekvenciju ponavljanja svake reči u okviru dokumenta i sastoji se od : reči, ID dokumenta i lokacije ili frekvencije reči unutar dokumenta. Sam indeks, možemo posmatrati kao skup parova oblika <word, document-id>, sortiranih po word polju. Za razliku od standardnih načina reprezentacije indeksa (B-stabla, koja omogućavaju upis i pretraživanje u), Lucene koristi nešto drugačiji pristup, koji umesto postojanja jednog indeksa, podrazumeva postojanje više indeks segmenata, čije se spajanje vrši periodično. Za svaki novi indeksirani dokument, Lucene kreira novi segment, pri čemu ubrzo vrši spajanje malih sa velikim segmentima – što održava mali ukupan broj segmenata. Radi optimizacije operacije pretraživanja, Lucene vrši spajanje svih segmenata u jedan, pri čemu se, radi izbegavanja mogućih konflikata konkurentnih čitanja i upisa u dati indeks, segmenti nikad ne modifikuju na mestu na kome se nalaze u indeksu, već se samo kreiraju novi. Pri spajanju, u novokreirani indeks, Lucene upisuje najnoviji indeks, dok vrši brisanje starijih, nakon što više nema aktivnih

procesa koji obavljaju čitanje datog indeksa. Ovaj pristup se jako dobro skalira u odnosu na veličinu indeksa, pri čemu omogućava fleksibilnost u proizvoljnom definisanju odnosa između brzine indeksiranja i brzine pretraživanja.

15.3 Rangiranje

Navedeni mehanizmi određivanja relevantnosti stranica u okviru indeksa, koje odgovaraju zadatom upitu, se pokazuju kao izuzetno efikasni u okviru klasičnog information “retrieval” domena. Međutim, u okviru problematike Web pretraživača, dati metodi se često pokazuju kao nedovoljno efikasni, upravo usled specifičnosti sadržaja na Web-u, koja se ogleda u postojanju strukture, indukovane hiperlinkovima između stranica. Usled ovoga, celokupan informacioni sadržaj o stranicama na Web-u se ne iscrpljuje isključivo posmatranjem sadržaja samih stranica, već zahteva i posmatranje informacija o odnosima između stranica.

15.3.1 PageRank

PageRank algoritam (koji se koristi u okviru Google pretraživača), ima za cilj dodeljivanje numeričke vrednosti u rasponu 0 do 1 (koja se naziva pagerank), svakom čvoru u Web Grafu, koja ukazuje na njegovu relevantnost, pri čemu data vrednost prvenstveno zavisi od same link strukture Web Grafa. Ukupna PageRank vrednost za svaku stranicu, se određuje za zadati upit, kao kombinacija pagerank vrednosti, određene strukturom Web Grafa i score vrednosti samog dokumenta, koja je određena samim sadržajem stranice (na način sličan opisanom određivanju score funkcije stranica u Lucene indeksu). Konačno, na osnovu dobijenih PageRank vrednosti, vrši se rankiranje liste rezultata dobijene u odnosu na zadati upit.

15.4 Clustering

Grupisanje podataka (clustering) predstavlja način obrade podataka, kojim se u samim podacima otkrivaju tzv. “grupe” (clusters) podataka koje pokazuju izvestan stepen “prirodne bliskosti”. Pod algoritmima za grupisanje podataka, podrazumevamo algoritme koji automatski otkrivaju grupe(“klastere”) u zadatom skupu podataka. Ovakav vid obrade podataka u praksi ima niz primena, posebno u oblasti “istraživanja podataka” (data mining).

Dva najčešća pristupa problematici grupisanja rezultata pretraživanja poklapaju se sa pristupima u okviru data mining metodologije (supervised/unsupervised learning). Kod supervised learning-a koristimo algoritme mašinskog učenja (naïve Bayes, SVM , Neuronske mreže), koji na osnovu zadatog skupa “ručno” klasifikovanih dokumenata (learning set), nakon procesa “učenja” (težinskih faktora kod naïve bayes, hiperprostora kod svm, koeficijena kod nn), vrše klasifikaciju stranica na osnovu njihovog sadržaja. Drugi pristup u okviru data mining metodologije predstavlja tzv. “unsupervised” learning, koji vrši klasifikaciju dokumenata bez prethodnog “znanja”, koristeći samo informacije prusutne u samim podacima (k-nearest neighbor). Ova metodologija se često koristi u “istraživanju podataka”, kada ne postoji prethodno definisani korpus podataka ili kada nismo sigurni šta tačno tražimo u podacima. Tome shodno, rezultate pretraživanja možemo posmatrati kao obične dokumente i primenjivati standardne metode mašinskog učenja za njihovo grupisanje (supervised – u odnosu na zadatak temu ili unsupervised – po sličnosti dokumenata). Ovakav pristup primenjuju neki komercijalni clustering pretraživači (Alexa, Clusty..).

16. Zaključak

Internet je jedan ogroman, nepresušan izvor podataka i ako ga naučimo koristiti, poboljšaćemo kvalitet života, a kako je on odlična alatka za učenje i izvor svežih informacija, pomoću njega možemo dobiti veoma bitne informacije. Internet je zajedničko ime za grupu informacionih resursa rasprostranjenih širom sveta. On se još naziva i mreža svih mreža jer su u nju povezane mreže obrazovnih i vladinih institucija kao i komercijalnih organizacija skoro svih zemalja sveta. Sve veća prisutnost Interneta u životu prosečnih i nadprosečnih korisnika kompjutera postala je opšte prihvaćena činjenica, samim tim postoji i potreba za korišćenjem pretraživača. Pretraživače često koristimo kao pomoć kod učenja i rada. Taj neumorivi pomoćnik pamti ogromnu količinu činjenica i ima pristup svom znanju ljudske vrste.

Kako je značaj pretraživača rastao tako su se razvijale i SEO tehnike. Nakon što je postao jasan marketinški značaj visokog kotiranja na pretraživačima, pojavile su se i kompanije koje pružaju SEO usluge, pa sam akronim SEO može stajati i za search engine optimizators. Nije dovoljno imati napravljenu Web stranicu, postaviti je na server i pustiti da vreme odradi svoje. Pre deset godina možda je i to bilo moguće, danas više ne. Web stranice nisu samo lep dizajn i kvalitetan sadržaj, one su „oružije“ kojim se na Internetu osvaja posao, zbog svega toga bitno je uraditi dobru optimizaciju sajta.

U okviru ovog master rada, prikazani su osnovni aspekti razvoja i implementacije pretraživačke mašine. Obrađeni su osnovni problemi, specifični za Web search problematiku, analizirane su njihove posledice i predstavljena odgovarajuća rešenja. Predstavljena su komercijalna rešenja (Google). Dodatno, je predstavljen razvoj sajta “www.elzing.rs”. Predstavljena je celokupna implementacija jednog Web sajta, na Internet, i njegova optimizacija na pretraživačima. Na kraju predstavljamo rezultate pretrage i posećenost sajta od njegovog puštanja u rad. Osnovni motiv samog rada jeste postavljanje, kako teorijske, tako i praktične osnove za dalji razvoj i istraživanje u oblasti Web pretraživanja.

17. Literatura

- [1] Lvsong Li, Jing Li
State Key Laboratory of Software Development Environment,
School of Computer Science and Engineering, Beihang University, Beijing,
China, 100191 {lilvsong, lijing}@nlsde.biaa.edu.cn
- [2] Bo Wang, Jinqiao Wang, Shi Chen, Ling-Yu Duan and Hanqing Lu
National Laboratory of Pattern Recognition, Institute of Automation, CAS, Beijing,
China {bwang, jqwang, schen, luhq} @nlpr.ai.ac.cn
- [3] Ian S. Burnett
RMIT University, Melbourne, Australia
- [4] Wu Oil, Luan Tianl, Bai Yanl, Wei Liyuanl, Li Yanhui
Software Technology Institute, Dalian Jiaotong University
- [5] A.K. Shanna, Neha Aggarwal, Neelam Duhan and Ranjna Gupta
Department of Computer Engineering, YMCA University of Science & Technology,
Faridabad, India, Department of Computer Science & Engineering, B.S.A. Institute of
Technology & Management, Faridabad, India
- [6] Mo Yunfeng
School of Information and Electronic Engineering
Zhejiang University of Science and Technology
Hangzhou, China 18462546@qq.com
- [7] College of Computer Science and Technology
Zhejiang University of Technology
Hangzhou, China 18462546@qq.com
- [8] <http://www.wikipedia.org>
- [9] <http://www.microsoft.com>
- [10] <http://www.oasis-open.org/>
- [11] <http://www.google.com>
- [12] <http://www.yahoo.com>